

US-09-020-116-4

Query Match 0.9%; Score 7; DB 3; Length 27;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563  
| | | | |  
Db 8 GKDSLSD 14

RESULT 25

US-09-608-902-4  
; Sequence 4, Application US/09608902  
; Patent No. 6353097  
; GENERAL INFORMATION:  
; APPLICANT: Vonakis, Becky  
; APPLICANT: Metzger, Henry  
; APPLICANT: Chen, Huaxian  
; TITLE OF INVENTION: SIGNAL TRANSDUCTION INHIBITORS OF  
; TITLE OF INVENTION: ALLERGIC REACTIONS  
; FILE REFERENCE: 14014.028502  
; CURRENT APPLICATION NUMBER: US/09/608,902  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 09/020,116  
; PRIOR FILING DATE: 1998-02-06  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: /No. 6353097e = synthetic cons

US-09-608-902-4

Query Match 0.9%; Score 7; DB 4; Length 27;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 557 GKDSLSD 563  
| | | | |  
Db 8 GKDSLSD 14

Search completed: October 1, 2004, 07:35:19  
Job time : 27 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:33:31 ; Search time 90 Seconds

(without alignments)

2846.135 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGLYQARTVKNRVS.....KLLALLKSNPSSVSKKIN 796

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1351062 seqs, 321799191 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	795	99.9	796	9	US-09-765-272-56
2	695	87.3	819	12	US-10-412-862-8
3	695	87.3	819	15	US-10-412-850-8
4	695	87.3	819	15	US-10-387-783-8
5	80	10.1	334	15	US-10-324-143-20
6	80	10.1	690	15	US-10-324-143-38
7	80	10.1	821	15	US-10-324-143-19
8	80	10.1	840	10	US-09-884-465A-7
9	80	10.1	840	15	US-10-324-143-8
10	60	7.5	819	15	US-10-412-862-10
11	60	7.5	819	15	US-10-324-143-32
12	60	7.5	819	15	US-10-412-850-10
13	60	7.5	819	15	US-10-387-783-10
14	60	7.5	826	10	US-09-769-787-194
15	60	7.5	838	10	US-09-884-465A-8

16	60	7.5	838	12	US-10-412-862-4	Sequence 4, Appli
17	60	7.5	838	15	US-10-324-143-9	Sequence 9, Appli
18	60	7.5	838	15	US-10-412-850-4	Sequence 4, Appli
19	60	7.5	838	15	US-10-387-783-4	Sequence 4, Appli
20	56	7.0	613	15	US-10-324-143-22	Sequence 22, Appli
21	51	6.4	763	9	US-09-765-272-66	Sequence 66, Appli
22	32	4.0	42	10	US-09-884-465A-12	Sequence 12, Appli
23	32	4.0	205	15	US-10-324-143-18	Sequence 18, Appli
24	32	4.0	447	9	US-09-765-272-182	Sequence 182, App
25	32	4.0	484	10	US-09-769-787-38	Sequence 38, Appli
26	32	4.0	484	10	US-09-769-744A-24	Sequence 24, Appli
27	32	4.0	484	12	US-10-412-862-6	Sequence 6, Appli
28	32	4.0	484	15	US-10-412-850-6	Sequence 6, Appli
29	32	4.0	484	15	US-10-387-783-6	Sequence 6, Appli
30	32	4.0	489	15	US-10-324-143-15	Sequence 15, Appli
31	32	4.0	509	15	US-10-324-143-16	Sequence 16, Appli
32	32	4.0	780	15	US-10-324-143-37	Sequence 37, Appli
33	32	4.0	840	10	US-09-884-465A-10	Sequence 10, Appli
34	32	4.0	1019	15	US-10-324-143-14	Sequence 14, Appli
35	32	4.0	1039	10	US-09-884-465A-6	Sequence 6, Appli
36	32	4.0	1039	15	US-10-324-143-7	Sequence 7, Appli
37	27	3.4	294	15	US-10-324-143-39	Sequence 39, Appli
38	27	3.4	473	15	US-10-324-143-36	Sequence 36, Appli
39	27	3.4	568	15	US-10-324-143-33	Sequence 33, Appli
40	27	3.4	612	15	US-10-324-143-44	Sequence 44, Appli
41	27	3.4	1139	10	US-09-884-465A-380	Sequence 380, App
42	27	3.4	1139	15	US-10-324-143-51	Sequence 51, Appli
43	27	3.4	1378	10	US-09-884-465A-378	Sequence 378, App
44	27	3.4	1378	15	US-10-324-143-49	Sequence 49, Appli
45	24	3.0	226	15	US-10-324-143-30	Sequence 30, Appli
46	24	3.0	428	15	US-10-324-143-27	Sequence 27, Appli
47	24	3.0	555	15	US-10-324-143-26	Sequence 26, Appli
48	24	3.0	999	10	US-09-884-465A-376	Sequence 376, App
49	24	3.0	999	10	US-09-884-465A-377	Sequence 377, App
50	24	3.0	999	15	US-10-324-143-47	Sequence 47, Appli
51	24	3.0	999	15	US-10-324-143-48	Sequence 48, Appli
52	24	3.0	1126	10	US-09-884-465A-383	Sequence 383, App
53	24	3.0	1126	15	US-10-324-143-54	Sequence 54, Appli
54	24	3.0	1238	10	US-09-884-465A-381	Sequence 381, App
55	24	3.0	1238	15	US-10-324-143-52	Sequence 52, Appli
56	24	3.0	1365	10	US-09-884-465A-382	Sequence 382, App
57	24	3.0	1365	15	US-10-324-143-53	Sequence 53, Appli
58	23	2.9	94	10	US-09-884-465A-11	Sequence 11, Appli
59	17	2.1	381	10	US-09-769-736-24	Sequence 24, Appli
60	17	2.1	793	10	US-09-252-088-15	Sequence 15, Appli
61	17	2.1	793	15	US-10-340-792-18	Sequence 18, Appli
62	17	2.1	822	10	US-09-769-736-18	Sequence 18, Appli
63	15	1.9	1058	15	US-10-324-143-46	Sequence 46, Appli
64	14	1.8	487	15	US-10-324-143-21	Sequence 21, Appli
65	14	1.8	487	15	US-10-324-143-35	Sequence 35, Appli
66	13	1.6	132	15	US-10-324-143-29	Sequence 29, Appli
67	12	1.5	203	15	US-10-324-143-31	Sequence 31, Appli
68	12	1.5	214	15	US-10-324-143-42	Sequence 42, Appli
69	12	1.5	342	15	US-10-324-143-40	Sequence 40, Appli
70	12	1.5	715	10	US-09-252-088-16	Sequence 16, Appli
71	12	1.5	715	15	US-10-340-792-16	Sequence 16, Appli
72	12	1.5	807	15	US-10-324-143-45	Sequence 45, Appli
73	12	1.5	913	10	US-09-884-465A-384	Sequence 384, App
74	12	1.5	913	15	US-10-324-143-55	Sequence 55, Appli
75	12	1.5	1152	10	US-09-884-465A-379	Sequence 379, App
76	12	1.5	1152	15	US-10-324-143-50	Sequence 50, Appli
77	11	1.4	485	10	US-09-769-736-72	Sequence 72, Appli
78	9	1.1	121	15	US-10-324-143-28	Sequence 28, Appli
79	9	1.1	124	12	US-10-424-599-256231	Sequence 256231,
80	9	1.1	272	15	US-09-884-465A-300	Sequence 300, App
81	9	1.1	272	15	US-10-324-143-85	Sequence 85, Appli
82	9	1.1	782	12	US-10-282-122A-44621	Sequence 44621, A
83	9	1.1	791	12	US-10-282-122A-67300	Sequence 67300, A
84	8	1.0	86	10	US-09-884-465A-22	Sequence 22, Appli
85	8	1.0	215	16	US-10-389-566-597	Sequence 597, App
86	8	1.0	257	15	US-10-369-493-9660	Sequence 9660, App
87	8	1.0	259	10	US-09-884-465A-374	Sequence 374, App
88	8	1.0	270	10	US-09-884-465A-302	Sequence 302, App

Sequence 87, Appl  
Sequence 258, App

8 1.0 270 15 US-10-324-143-87  
90 1.0 272 10 US-09-884-465A-258

# ALIGNMENTS

RESULT 1  
US-09-765-272-56  
; Sequence 56, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 56:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 796 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-765-272-56  
Query Match 99.9%; Score 795; DB 9; Length 796;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

Db 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

QY 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIINEVKGKGVYKVDGKYVYLKDA 120

Db 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIINEVKGKGVYKVDGKYVYLKDA 120

QY 121 AHADNVRTKEENRQKQESQREHGGTTPNDGVALARSQGRYTTDDGVIFNASDIIEDT 180

Db 121 AHADNVRTKEENRQKQESQREHGGTTPNDGVALARSQGRYTTDDGVIFNASDIIEDT 180

QY 181 GDAYIVPGHDHYHYIPKNELSAELAAEAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240

Db 181 GDAYIVPGHDHYHYIPKNELSAELAAEAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240

QY 241 VSNPGTTNTNTNNSNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITST 300

Db 241 VSNPGTTNTNTNNSNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITST 300

Db 241 VSNPGTTNTNTNNSNTNSQASQSDNDISLLKQLKPLSORHVESDGLVFPDPAITST 300

QY 301 ARGVAVPGHDHYHYIPYSQMSLEERIIARIPLRYSRHHWVPDSRPEQSPQPTPEPSPG 360

Db 301 ARGVAVPGHDHYHYIPYSQMSLEERIIARIPLRYSRHHWVPDSRPEQSPQPTPEPSPG 360

QY 361 POPAPNLKIDSNSSLSVQLVRKVGSGYVFEKGIISRYVFAKDLPSSETVKNLESKLSKORS 420

Db 361 POPAPNLKIDSNSSLSVQLVRKVGSGYVFEKGIISRYVFAKDLPSSETVKNLESKLSKORS 420

QY 421 VSHTLTAKKENVAPRDQBFYDKAYNLLTEAHKALFXNKGSRNSDFQALDKLLERLNDESTN 480

Db 421 VSHTLTAKKENVAPRDQBFYDKAYNLLTEAHKALFXNKGSRNSDFQALDKLLERLNDESTN 480

QY 481 KEKLVDDLLAFAPITTHPERLGKPNQJTEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540

Db 481 KEKLVDDLLAFAPITTHPERLGKPNQJTEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540

QY 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQVYTKEGILPPSPDADVKANPTGDSAAAIY 600

Db 541 EGDAYVTPHMGHSHWIGKDSLSDEKVAQAQVYTKEGILPPSPDADVKANPTGDSAAAIY 600

QY 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWEDDHTYKAPNGYTLED 660

Db 601 NRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWEDDHTYKAPNGYTLED 660

QY 661 LPATIKYVVEHPDERPHSNDGWNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720

Db 661 LPATIKYVVEHPDERPHSNDGWNASEHVLGKDHSEDPNKNFKADEPVEETPAEPEVP 720

QY 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETTLAQLRNNLTLOIMDNNSIMAEAEKLLA 780

Db 721 QVETEKVEAQLKEAEVLLAKVTDSLSKANATETTLAQLRNNLTLOIMDNNSIMAEAEKLLA 780

QY 781 LLKGSNPSSVSKEKIN 796

Db 781 LLKGSNPSSVSKEKIN 796

RESULT 2  
US-10-412-862-8  
; Sequence 8, Application US/10412862  
; Publication No. US20040052781A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
; TITLE OF INVENTION: Motifs  
; FILE REFERENCE: 469201-685  
; CURRENT APPLICATION NUMBER: US/10/412,862  
; CURRENT FILING DATE: 2003-04-14  
; PRIOR APPLICATION NUMBER: 09/468,656  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/113,048  
; PRIOR FILING DATE: 1998-12-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 8  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-412-862-8

Query Match 87.3%; Score 695; DB 12; Length 819;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 60

Db 21 SYELGLYQARTVKNRRVSYIDGKQATKTENLTDPDESVKREGINAEQIVIKITDQGYVT 80

QY 61 SHGDHYHYNGKVPDVAIISEBLLMKDPNYKLDKEDIINEVKGKGVYKVDGKYVYLKDA 120



```
Db      81 SHGDHYYNGKVPYDAIISEELMKDPNKLKDEDIVNEVKGYIVKVDGKYVYLKDA 140
QY      121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNASDIIEDT 180
Db      141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNASDIIEDT 200
QY      181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRQNSONTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRQNSONTSRTNWVPS 260
QY      241 VSNPGTTNTNTSNNTNSOASQNSDIDSLKQYKLPISORHVESDGLVDFDPAQITTSRT 300
Db      261 VSNPGTTNTNTSNNTNSOASQNSDIDSLKQYKLPISORHVESDGLVDFDPAQITTSRT 320
QY      301 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSG 360
Db      321 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSG 380
QY      361 PQAPNLKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQES 420
Db      381 PQAPNLKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQES 440
QY      421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY      481 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY      541 EGDAYVTPHMGHSHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADYKANPTGDSAAIY 600
Db      561 EGDAYVTPHMGHSHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADYKANPTGDSAAIY 620
QY      601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTTKAPNGYTTLED 660
Db      621 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTTKAPNGYTTLED 680
QY      661 LFATIKYVVEHPDHPHSDNGWGNASHVLGKDHSEDPNKNFKADBEPEETPAEPEVP 720
Db      681 LFATIKYVVEHPDHPHSDNGWGNASHVLGKDHSEDPNKNFKADBEPEETPAEPEVP 740
QY      721 QVETEKVEAQLKEAEVLLAKVTSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
Db      741 QVETEKVEAQLKEAEVLLAKVTSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY      781 LLKGSNPSSVSKEKIN 796
Db      801 LLKGSNPSSVSKEKIN 816
```

## RESULT 3

```
US-10-412-850-8
; Sequence 8, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
```

```
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-8
```

```
Query Match      87.3%; Score 695; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 SYELGLYQARTVVENNRVSYIDOKAQTKQTKTENITPDEVSKREGINABQIVIKITDQGYVT 60
Db      21 SYELGLYQARTVVENNRVSYIDOKAQTKQTKTENITPDEVSKREGINABQIVIKITDQGYVT 80
QY      61 SHGDHYYNGKVPYDAIISEELMKDPNKLKDEDIVNEVKGYIVKVDGKYVYLKDA 120
Db      81 SHGDHYYNGKVPYDAIISEELMKDPNKLKDEDIVNEVKGYIVKVDGKYVYLKDA 140
QY      121 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNASDIIEDT 180
Db      141 AHADNVRTKEEINRQKQHSQHREGGTPRNDGAVALARSQRYTTDDGYIFNASDIIEDT 200
QY      181 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRQNSONTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIIPKNELASASELAFLSGRGNLSNRYRQNSONTSRTNWVPS 260
QY      241 VSNPGTTNTNTSNNTNSOASQNSDIDSLKQYKLPISORHVESDGLVDFDPAQITTSRT 300
Db      261 VSNPGTTNTNTSNNTNSOASQNSDIDSLKQYKLPISORHVESDGLVDFDPAQITTSRT 320
QY      301 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSG 360
Db      321 ARGVAVPHGDHYHYIIPYSQMSLEERARIIPLYRSNHWVPSRPEQSPQPTPEPSG 380
QY      361 PQAPNLKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQES 420
Db      381 PQAPNLKIDNSLSVQLVRKVGEGYVFEKGISRYVFAKDLPSSETVKNLESKLSQES 440
QY      421 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db      441 VSHLTAKKENVAPRQDEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500
QY      481 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 540
Db      501 KEKLVDDLAFAPITHPERLGKPNQSIYEYDEVRIAQLADKYTTSDGYIFDEHDIISD 560
QY      541 EGDAYVTPHMGHSHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADYKANPTGDSAAIY 600
Db      561 EGDAYVTPHMGHSHWIGKDSLSKKEKVAQAQYTKGKILPPSPDADYKANPTGDSAAIY 620
QY      601 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTTKAPNGYTTLED 660
Db      621 NRVKGEKRIPLVRLPYMVEHTVEKNGNLIIPHKHVHNIKFAWFDHDTTKAPNGYTTLED 680
QY      661 LFATIKYVVEHPDHPHSDNGWGNASHVLGKDHSEDPNKNFKADBEPEETPAEPEVP 720
Db      681 LFATIKYVVEHPDHPHSDNGWGNASHVLGKDHSEDPNKNFKADBEPEETPAEPEVP 740
QY      721 QVETEKVEAQLKEAEVLLAKVTSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
Db      741 QVETEKVEAQLKEAEVLLAKVTSSKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 800
QY      781 LLKGSNPSSVSKEKIN 796
Db      801 LLKGSNPSSVSKEKIN 816
```

## RESULT 4

```
US-10-387-783-8
; Sequence 8, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
```

; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 8
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-8

Query Match 87.3%; Score 695; DB 15; Length 819;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 60
Db 21 SYELGLYQARTVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVT 80

QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKEDIVNEVKGYYIKVDGKYYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISELLMKDPNYKDKEDIVNEVKGYYIKVDGKYYVYLKDA 140

QY 121 AHADNVRKEENRQKHSQHREGTTPNDGAVALARSQRYTTDDGVIFNASDIIDT 180
Db 141 AHADNVRKEENRQKHSQHREGTTPNDGAVALARSQRYTTDDGVIFNASDIIDT 200

QY 181 GDVIVPHGDHVIYPKNELSASELAFAFUSGRGNSLSTYRQNSDNTSRNNWVPS 240
Db 201 GDVIVPHGDHVIYPKNELSASELAFAFUSGRGNSLSTYRQNSDNTSRNNWVPS 260

QY 241 VSNPGTTNTNNSNTNSQASQNSIDSLKQLYKPLSQRHVSVDGLVFPDPAQITST 300
Db 261 VSNPGTTNTNNSNTNSQASQNSIDSLKQLYKPLSQRHVSVDGLVFPDPAQITST 320

QY 301 ARGVAVPHGDHVIYPYQMSLEBRARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 360
Db 321 ARGVAVPHGDHVIYPYQMSLEBRARIIPLYRSNHWVPSRPEQSPQPTPEPSPG 380

QY 361 POPAPNLKIDNSLSVQLVRKVGYYVEEKGISRYVFAKDLPSSETVKNLESKLSKQES 420
Db 381 POPAPNLKIDNSLSVQLVRKVGYYVEEKGISRYVFAKDLPSSETVKNLESKLSKQES 440

QY 421 VSHTLTAKKENVAPRDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
Db 441 VSHTLTAKKENVAPRDQFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 500

QY 481 KEKLVDLLAFAPITHEPRLGKPNQSYTEDEVRIAQLADKYTTSDGYIPDEHDIISD 540
Db 501 KEKLVDLLAFAPITHEPRLGKPNQSYTEDEVRIAQLADKYTTSDGYIPDEHDIISD 560

QY 541 EGDVYVTPMGHSHWIKGDSLSDEKVAQAAYTEKEGILPPSPDADVKANPTGDSAAAIY 600
Db 561 EGDVYVTPMGHSHWIKGDSLSDEKVAQAAYTEKEGILPPSPDADVKANPTGDSAAAIY 620

QY 601 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWPDHHTYKAPNGYTTLED 660
Db 621 NRVGKGRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNIKPAWPDHHTYKAPNGYTTLED 680

QY 661 LPATIKYVVEHPDERPHSNDGNGNASEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 720
Db 681 LPATIKYVVEHPDERPHSNDGNGNASEHVLGKDHSEDPNKNFKADEEPVEETPAEPVP 740

QY 721 QVETEKVBAQKEAEVLLAKVTDSSLKANATETLAGLNNLTQLQMDNNSIMAEAEKLLA 780
Db 741 QVETEKVBAQKEAEVLLAKVTDSSLKANATETLAGLNNLTQLQMDNNSIMAEAEKLLA 800

QY 781 LLKGSNPSSVSKEKIN 796
Db 801 LLKGSNPSSVSKEKIN 816

RESULT 5
US-10-324-143-20
; Sequence 20, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLEAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 20
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-20

Query Match 10.1%; Score 80; DB 15; Length 334;
Best Local Similarity 100.0%; Pred. No. 3.9e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 70
Db 12 TVKENNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYVTSBGDHYHYN 71

QY 71 GKVPYDAIISELLMKDPNY 90
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 6
US-10-324-143-38
; Sequence 38, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLEAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 38
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-38

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match      10.1%; Score 80; DB 15; Length 690;
Best Local Similarity 100.0%; Pred. No. 7.8e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 12 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 71

Qy 71 GKVPYDAIISELLMKDPNY 90
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 7
US-10-324-143-19
; Sequence 19, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-19

Query Match      10.1%; Score 80; DB 15; Length 821;
Best Local Similarity 100.0%; Pred. No. 9.1e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 12 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 71

Qy 71 GKVPYDAIISELLMKDPNY 90
Db 72 GKVPYDAIISELLMKDPNY 91

RESULT 8
US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

Query Match      10.1%; Score 80; DB 10; Length 840;
Best Local Similarity 100.0%; Pred. No. 9.3e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 31 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 90

Qy 71 GKVPYDAIISELLMKDPNY 90
Db 91 GKVPYDAIISELLMKDPNY 110

RESULT 9
US-10-324-143-8
; Sequence 8, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-8

Query Match      10.1%; Score 80; DB 15; Length 840;
Best Local Similarity 100.0%; Pred. No. 9.3e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 70
Db 31 TVKENNRVSYIDGKQATKTKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYN 90

Qy 71 GKVPYDAIISELLMKDPNY 90
Db 91 GKVPYDAIISELLMKDPNY 110

RESULT 10
US-10-412-862-10
; Sequence 10, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
```

```
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; TYPE: PRT
; LENGTH: 819
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-10

Query Match
Best Local Similarity 7.5%; Score 60; DB 12; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 90
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 11
US-10-324-143-32
; Sequence 32, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEF
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; TYPE: PRT
; LENGTH: 819
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-32

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 90
DB 33 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 92

RESULT 12
US-10-412-850-10
; Sequence 10, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
```

```
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; TYPE: PRT
; LENGTH: 819
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-10

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 90
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 13
US-10-387-783-10
; Sequence 10, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-10

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 90
DB 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDAIISELLMKDPNY 111

RESULT 14
US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P211290
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 194
```

```
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-194

Query Match      7.5%  Score 60;  DB 10;  Length 826;
Best Local Similarity 100.0%;  Pred. No. 6e-50;
Matches 60;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 15
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match      7.5%  Score 60;  DB 10;  Length 838;
Best Local Similarity 100.0%;  Pred. No. 6.1e-50;
Matches 60;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 16
US-10-412-862-4
; Sequence 4, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-4
```

```
Query Match      7.5%  Score 60;  DB 12;  Length 838;
Best Local Similarity 100.0%;  Pred. No. 6.1e-50;
Matches 60;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 17
US-10-324-143-9
; Sequence 9, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELLETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-9

Query Match      7.5%  Score 60;  DB 15;  Length 838;
Best Local Similarity 100.0%;  Pred. No. 6.1e-50;
Matches 60;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 90
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSHGHDHYHYNGKVPYDAIISEELMKDPNY 111

RESULT 18
US-10-412-850-4
; Sequence 4, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-4

Query Match      7.5%  Score 60;  DB 15;  Length 838;
Best Local Similarity 100.0%;  Pred. No. 6.1e-50;
Matches 60;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

Matches	60;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

Qy	31	ENLTPDEVSKREGINAEQIVIKITDQGYVTS	SHGDHYHYNGKVPYDAIISBELLMKDPNY	90
D6	52	ENLTPDEVSKREGINAEQIVIKITDQGYVTS	SHGDHYHYNGKVPYDAIISBELLMKDPNY	111

```

RESULT 19
US-10-387-783-4
; Sequence 4, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-4

```

Qy 31 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 90  
|||  
Db 52 ENLTPDEVSKREGINAEQIVIKITDQGYVTSBGDHYHYNGKVPYDAIISELLMKDPNY 111

```

RESULT 20
US-10-324-143-22
; Sequence 22, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OUELETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 55190-56
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-22

```

Qy 234 RTWVPSVSNPGTTNTNTSNNSTNSQASQSNIDSLKQLYKPLSQRVESDGL 289  
|||  
Db 27 RTWVPSVSNPGTTNTNTSNNSTNSQASQSNIDSLKQLYKPLSQRVESDGL 82

```

RESULT 21
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Hp Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

```

Qy 160 QGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAEA 210  
|||  
Db 159 QGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAEA 209  
|||

```

RESULT 22
US-09-884-465A-12
/ Sequence 12, Application US/09884465A
/ Publication No. US20030077293A1
/ GENERAL INFORMATION:
/ APPLICANT: Shire Biochem, Inc.
/ APPLICANT: Hamel, Josee
/ APPLICANT: Brodeur, Bernard
/ APPLICANT: Martin, Denis
/ APPLICANT: Charland, Nathalie
/ APPLICANT: Ouellet, Catherine
/ TITLE OF INVENTION: Streptococcus An
/ FILE REFERENCE: 055190-0044
/ CURRENT APPLICATION NUMBER: US/09/884
/ CURRENT FILING DATE: 2001-06-20

```

```
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-12

Query Match          4.0%; Score 32; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 2e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 77
Db 4 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 35

RESULT 23
US-10-324-143-18
; Sequence 18, Application US/10324143
; Publication No. US20030232976A1
; GENERAL INFORMATION:
; APPLICANT: HAMEL, JOSEE
; APPLICANT: CHARLAND, NATHALIE
; APPLICANT: BRODEUR, BERNARD R.
; APPLICANT: MARTIN, DENIS
; APPLICANT: BLAIS, NORMAND
; APPLICANT: OULETTE, CATHERINE
; TITLE OF INVENTION: STREPTOCOCCUS ANTIGENS
; CURRENT APPLICATION NUMBER: US/10/324,143
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/341,252
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: modified Streptococcus pneumoniae protein sequence
US-10-324-143-18

Query Match          4.0%; Score 32; DB 15; Length 205;
Best Local Similarity 100.0%; Pred. No. 9e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 77
Db 47 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 78

RESULT 24
US-09-765-272-182
; Sequence 182, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182

Query Match          4.0%; Score 32; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 77
Db 43 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 74

RESULT 25
US-09-769-787-38
; Sequence 38, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-38

Query Match          4.0%; Score 32; DB 10; Length 484;
Best Local Similarity 100.0%; Pred. No. 2e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 77
Db 67 AEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 98

Search completed: October 1, 2004, 07:41:06
Job time : 92 secs
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 1, 2004, 07:35:26 ; Search time 7905 Seconds  
(without alignments)  
4364.460 Million cell updates/sec

Title: US-09-765-271-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKENNRVSY.....KLALLKGSNPFSSVSKKIN 796

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool\_h/US09765271/runat\_30092004.113753.12277/app\_query.fasta\_1.967  
-DB=GenEmbl -QMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=90  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL  
-USER=US09765271.qcgn 1.1 5239 @runat\_30092004.113753.12277 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.in:\*  
18: em.hum:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg.hum:\*  
31: em.htg.inv:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rtd:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.sy:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	4163	100.0	2389	6	AR120265 Sequence
2	4163	100.0	2389	6	AR340956 Sequence
3	4163	100.0	2389	6	BD063274 Streptoco
4	4163	100.0	2406	6	AX569139 Sequence
5	4163	100.0	2451	6	BD268048 Streptoco
6	4163	100.0	2451	6	AR344447 Sequence
7	4163	100.0	2541	1	AF291695 Sequence
8	4163	100.0	8195	6	AR218862 Streptoco
9	4163	100.0	8195	6	BD003774 Polynucle
c 10	4163	100.0	10256	1	AE007418 Streptoco
c 11	4163	100.0	349980	6	AX571763 Sequence
c 12	4159	99.9	10320	1	AE008479 Streptoco
13	4112	97.7	20035	2	SPNEU1915 Streptoco
14	3236	77.7	2535	1	AF340221 Streptoco
15	3218	77.3	2523	6	BD263585 Novel str
16	3218	77.3	2523	6	AX343072 Sequence
17	3218	77.3	2647	6	BD263590 Novel str
18	3218	77.3	2647	6	AX343073 Sequence
19	2812	67.5	2478	6	BD229972 Human com
20	2797	67.2	2457	6	AX569137 Sequence
21	2795	67.1	2481	6	BD264091 Streptoco
22	2786	66.9	2457	1	AF318954 Streptoco
23	2786	66.9	2531	6	BD268049 Streptoco
24	2786	66.9	2531	6	AR344448 Sequence
25	2781	66.8	232807	2	SPNEU1901 Streptoco
26	2772	66.6	2517	1	AF318955 Streptoco
27	2772	66.6	2517	6	AX568778 Sequence
28	2772	66.6	2531	6	BD268046 Streptoco
29	2772	66.6	2531	6	AR344445 Sequence
30	2772	66.6	11931	1	AE007403 Streptoco
31	2772	66.6	349980	6	AX571762 Sequence
32	2750	66.0	2639	6	BD263591 Novel str
33	2750	66.0	2639	6	AX343074 Sequence
34	2733	65.6	12372	1	AE008464 Streptoco
35	2649	63.6	2290	6	AR120270 Sequence
36	2649	63.6	2290	6	AR340961 Sequence
37	2649	63.6	2290	6	BD063279 Streptoco
38	2431	58.4	2166	12	AF340222 Synthetic
c 39	2382	57.2	75874	2	SPNEU1907 Streptoco
40	1774	42.6	2359	6	AR219011 Sequence
41	1774	42.6	2359	6	BD003923 Polynucle
42	1591	38.2	3171	6	BD263631 Novel str
43	1298	31.2	75248	6	AX602204 Sequence
c 44	1298	31.2	98050	1	SAG766854 Streptoco
45	1247	29.9	3120	6	BD263584 Novel str
46	1247	29.9	3120	6	AX343070 Sequence
47	1247	29.9	5048	6	BD263589 Novel str
48	1247	29.9	5048	6	AX343071 Sequence
49	1246	29.9	3117	6	AX568780 Sequence
50	1246	29.9	3120	1	AF318956 Streptoco

C 51	1222.5	29.4	6867	6	AR218960	Sequence
C 52	1222.5	29.4	6867	6	BD003872	Polynucle
C 53	1216.5	29.2	21397	1	AE014279	Streptoco
C 54	1204	28.9	702	2	SPNEU1929	Streptoco
C 55	1203.5	28.9	1455	6	BD263587	Novel str
C 56	1203.5	28.9	1455	6	BD264131	Streptoco
C 57	1203.5	28.9	1455	6	BD268047	Streptoco
C 58	1203.5	28.9	1455	6	AR344446	Sequence
C 59	1203.5	28.9	1455	6	BD223895	Nucleic a
C 60	1187.5	28.5	1342	6	AR120328	Sequence
C 61	1187.5	28.5	1342	6	AR341019	Sequence
C 62	1187.5	28.5	1342	6	BD063337	Streptoco
C 63	995	23.9	1446	6	AX608407	Sequence
C 64	988	23.7	973	6	AR219123	Sequence
C 65	988	23.7	973	6	BD004035	Polynucle
C 66	966.5	23.2	2528	6	BD263592	Novel str
C 67	966.5	23.2	2528	6	AX343078	Sequence
C 68	950	22.8	2478	6	AX088372	Sequence
C 69	950	22.8	10825	1	AE006623	Streptoco
C 70	947	22.7	10618	1	AE010110	Streptoco
C 71	946	22.7	2469	6	BD263632	Novel str
C 72	946	22.7	2469	6	AX607520	Sequence
C 73	946	22.7	5215	6	BD136831	Group B s
C 74	946	22.7	11617	6	AX602163	Sequence
C 75	946	22.7	199050	1	AL766850	Streptoco
C 76	945	22.7	2478	1	AB073859	Streptoco
C 77	941	22.6	2472	6	BD263633	Novel str
C 78	941	22.6	50354	1	AE014169	Streptoco
C 79	941	22.6	323825	1	AP005146	Streptoco
C 80	937	22.5	21414	1	AE014248	Streptoco
C 81	934	22.4	2469	6	AX088376	Sequence
C 82	934	22.4	3501	1	AF062533	Streptoco
C 83	929	22.3	83	6	BD263929	Streptoco
C 84	841	20.2	1146	6	BD263932	Nucleic a
C 85	833	20.0	10312	1	AE007404	Streptoco
C 86	770.5	18.5	1684	6	AR219026	Sequence
C 87	770.5	18.5	1684	6	BD003938	Polynucle
C 88	679.5	16.3	1910	1	SAG290952	Streptoco
C 89	650	15.6	2721	6	AX799438	Sequence
C 90	644	15.5	816	6	AX799436	Sequence

  

RESULT 1	AR120265	2389 bp	DNA	linear	PAT 16-MAY-2001
LOCUS	Sequence	55 from patent US 6159469.			
DEFINITION	AR120265				
ACCESSION	AR120265				
VERSION	AR120265.1	GI:14103841			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2389)				
AUTHORS	Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R., and Rosen, C.A.				
TITLE	Streptococcus pneumoniae antigens and vaccines				
JOURNAL	Patent: US 6159469-A 55 12-DEC-2000;				
FEATURES	Location/Qualifiers				
source	1..2389				
	/organism="unknown"				
	/mol_type="unassigned DNA"				

  

Alignment Scores:	7.42e-230	Length:	2389
Pred. No.:	4163.00	Matches:	796
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	99.95%	Indels:	0
Query Match:	6	Gaps:	0
DB:			

US-09-765-271-56 (1-796) x AR120265 (1-2389)

1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20  
 2 TCTTACGAGTTGGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATPATTCGTTTCTTAT 61  
 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 62 ATAGATGAAACCAAGCCGACCAAAACGAGAAATTTGACTCTCTGATGAGGTAGCAAG 121  
 41 ArgGluGlyLysAlaAsnAlaGluGlnIleValLysLysIleThrAspGlnGlyTyrValThr 60  
 122 CQTGAAGGAATCAATCTGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 181  
 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 182 TCACATGGCGACCATCTATCATTTATCAATGGTAGGTTCTTATGACGCTATCATCAGT 241  
 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 242 GAAGAAATTAATCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 301  
 101 ValLysGlyLysTyrValLysValLysValAspGlyLysTyrTyrValTyrLysAspAla 120  
 302 GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAAGGATGCT 361  
 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 362 GCCCAGCGGATTAAGTCGTCACAAAGAGGAAATCAATCGCAAAAACCAAGGCATAGT 421  
 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 422 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGTTGCTTGGCCTTGGCAAGTTCGCAA 481  
 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 482 GGACGCTATCTACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATCT 541  
 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 542 GGTGATGCTTATATGTTCTCTCATGGAGATCATACCATTACATTACCTTAAGAATGAGTTA 601  
 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 602 TCAGCTAGCAGTTGGCTGCTGCAGAAAGCTTCTCTATCTGTCGAGGAAATCTGTCAAAT 661  
 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 662 TCAAGAACCTATCGCCGACAAATAGCGATAACACTTCAAGAACAAACTGGGTACCTTCT 721  
 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 722 GTAAGCAATCCAGGAATCAATACTAACACAGCAACACACAGCAACACTAACAGTCAA 781  
 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrIleLysLeuProLeuSer 280  
 782 GCAAGTCAAAGTAATGACATTGATGCTCTTGAAACAGCTCTACAAACTGCGCTTTGAGT 841  
 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 842 CAACGACATGTGAATCTGATGGCTTGTCTTTGATCCAGCACAAAATCACAAGTCGACACA 901  
 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 902 GCTAGAGGTGTTGACAGTCCACAGGAGATCAATACCACCTTCATCCCTTACTCTCAAAATG 961  
 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 962 TCTGAATTTGAGAGACGAATCGCTCGCTATTAATTCCTCTTATCGTTTCTCAAACTATGG 1021  
 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 1022 GTACCAAGATTCAAGGCCGACGACCAACCAAGTCCACACCGACTCCGGACCTAGTCCAGGC 1081

```
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerClnleuVal 380
Db 1082 CCGNACCTGACCCNAACTTAAATAGACTCAAAATCTCTTTGGTAGTCAGCTGGTA 1141
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1142 CGAAAAGTTGGGAAGGATATGTAATTCGAAGAAAAGGCATCTCTCGTTATGCTCTTGGG 1201
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1202 AAAGATTTTACCATCTCGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAAGAGT 1261
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProAspGluGlnGluPheTyr 440
Db 1262 GTTTCACACATTTAACTGCTTAAAGAAATGTTGCTCTCGTGACCAAGATTTTAT 1321
QY 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCGT 1381
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1382 AATCTGATTTTCCAGGCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAAT 1441
QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1442 AAAGAAAATTTGTAGATGATTTATGTCATCTTAGCACCANTTACCATCCAGAGCGA 1501
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1502 CTTGGCAAAACCAATCTCAAAATGAGTATCTGAAGACGAAGTTGCTGATTTGCTCAATTA 1561
QY 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1562 GCTGATAGTATACAAGCTCAGATGGTTACATTTTGTGATGAACATGATATATATCAGTGAT 1621
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
Db 1622 GAAGAGATGCATATGTAACGCTCATATGGCCATAGTCTACTGGATTGGAAAAGATAGC 1681
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 1682 CTTTCTGATAGGAAAAAGTTGCAGCTCAAGCTTATCTAAAGAAAAGGTTATCTTACCT 1741
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 1742 CCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTAC 1801
QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 1802 AATCGTGTGAAGGGGAAAACGAATTCCTGCTGCTGCTTCCATATATGTTGAGCAT 1861
QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
Db 1862 ACAGTTGAGGTTAAAAACGGTAATTTGATTTCTCTCATAGGATCATTTACCATATAT 1921
QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 1922 AAATTTGCTGTTTGAATGATCACAATACCAAGCTCCAAATGCTATACCTTGGGAAGAT 1981
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAspAsp 680
Db 1982 TTGTTTGGCAGATTAAGTACTAGTAGAACCCCTGACGAACCTCCACATTTCAATGAT 2041
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 2042 GGATGGGGCAATGCCAGTGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAAT 2101
QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
Db 2102 AAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161
```

```
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 2162 CAAGTAGAGACTGAAAAAGTAGAACCCCAACTCAAGAAGCAGAAAGTTTGTCTTCGAAA 2221
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 2222 GTNACGGATTCTAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACGAATTAAT 2281
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
Db 2282 TTGACTCTTCAATTTATGGATAACAATAGTATCATGGCAGAACGAGAAAAATTTACTTGG 2341
QY 781 LeuLysLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2342 TTGTTAAAGGAAGTATCTCTTCTGTAAGTAGGAAAAATAAAC 2389

RESULT 2
AR340956 2389 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION Sequence 55 from patent US 6573082.
ACCESSION AR340956
VERSION AR340956.1 GI:33732935
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE Unclassified.
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Rosen, C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6573082-A 55 03-JUN-2003;
FEATURES
    Location/Qualifiers
        source
            1..2389
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 7,42e-230 Length: 2389
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AR340956 (1-2389)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnAtgValSerTyr 20
Db 2 TCTTACGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATATCTGTTTCTTAT 61
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 62 ATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAAG 121
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 122 CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 181
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 182 TCATGCGGACCACTATCATATTATCANTGGTAAGGTTCTCTTATGACGCTATCATCAGT 241
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 242 GAAGAATTTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTTGTAATGAG 301
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 302 GTCAAGGGTGGATGTTATCAAGGTAGTAGGAAAAATCTATGTTTACCTTAAGGATGCT 361
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
```

Db 362 GCCACCGGATACCGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 421  
 Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 422 CAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGTGCTTGGCCTTGCACGTTCCGCAA 481  
 Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 482 GGACGCTATACACAGATGATGGTATATCTTTAAATGCTTCGATATCATAGAGATACT 541  
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 542 GGTGATGCTTATATCGTTCCTCATGGAGATCATTTACCATTTACCTTAAGAAATGAGTTA 601  
 Qy 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 602 TCAGCTAGCGAGTTGGCTGCTCGAAGAGCTTCTATCTGCTGGTCGAGGAAATCTGCAAAAT 661  
 Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 662 TCAAGAACCTATCGCGACAAATAGCGATACACTTCAAGAACCAAACTGGGTACCTTCT 721  
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 722 GTAAGCAATCCAGGAATCAAAATACTAAACAAGCAACAACAGCAACACTAACAGTCAA 781  
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 782 GCAAGTCAAAAGTAATGACATTCATGATGCTCTTGAACAGCTCTACAAACTGGCTTTGAGT 841  
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 842 CAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAGTCGAACA 901  
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 902 GCTAGAGTGTTCAGTGCACAGGAGATCATTAACCTTCATCCCTTACTCTCAAAATG 961  
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 962 TCTGAATGGAGAAACGAATCGCTCGTATATTTCCCTTCGTATCTGTTATCTTCAACCAATGG 1021  
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProGlnProGlnProSerProGly 360  
 Db 1022 GTACCAAGTTCAGGCGAGACCAACCAAGTCCACACCGACCTCCGGAACCTAGTCAGGC 1081  
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerIleuValSerGlnLeuVal 380  
 Db 1082 CGGCAACCTGCACAAATCTTAAATATAGACTCAAAATTCCTTTGTTAGTCAGCTGGTA 1141  
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1142 CGAAAAGTGGGGAAGATGATGTTATCGAAGAAAGGGATCTCTCTGTTATGCTTTGGC 1201  
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerGlnGluSer 420  
 Db 1202 AAAGATTTACCATCTGAACTGTAAATCTTTGAAGCAAGTTCATCAAAACAAGAGAT 1261  
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1262 GTTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCACCAAGAAATTTAT 1321  
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1322 GATAAGCATATATCTGTAACTGAGGCTCATAGGCTTGTGTTGNAATAAGGTCGT 1381  
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1382 AATTCCTGATTCGAAGCCTTAGACAAATATTAGAACGCTTGAATGAATGAATCGACTAAT 1441  
 Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1442 AAAGAAAAATTTGGTAGATGATTTATTGGCATTCCTAGCACCAATTAACCATCCAGAGCGA 1501

Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1502 CTTGGCAACCAAAATCTCAAAATTTAGTATACTGAAGCAAGTTCGTTAGTCTCAATTA 1561  
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1562 GCTGATAAGTATACAACTGATGTTTACATTTTGTATGAAATGATATATATCATAGTAT 1621  
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 1622 GAAGGAGATGATATGTAAACGCTCATATGGGCCATAGTCACTGGATTTGGAAGATAGC 1681  
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1682 CTTTCTGATAGGAAAAAGTTGCCAGTCAAGCCTATCTATAAGAAAAAGGTATCTACCT 1741  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1742 CCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCCAGCAGCTATTTAC 1801  
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1802 AATCGTGTGAAGAGGAAAAACGAATTCACCTCGTTCGACTTCCATATATGTTGAGCAT 1861  
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1862 ACAGTTGAGTTAAAAACGGTAATTTGATTTATCTCTCATAGGATCATTTACCATAATTT 1921  
 Qy 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1922 AAAATTTGCTTGGTTTGTATGATCACACATACAAAGCTCCAAATGGCTATATACCTTGAAGAT 1981  
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 1982 TTGTTTGGCAGCATTAAGTACTACGTAGAAACCCCTGACGACGCTCCACATTTCTATGAT 2041  
 Qy 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2042 GGATGGGCAATGCCAGTGAGCATGTGTGTAGCAAGAAAGACCAACAGTGAAGATCCAAAT 2101  
 Qy 701 LysAsnPheLysAlaAspGluLeuProValGluGluThrProAlaGluProGluValPro 720  
 Db 2102 AAGAACTTCAAAACGCGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161  
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2162 CAAGTAGACACTGAAAAGTAGAGCCCAACTCAAGAAACAGAGTTTGTTCGCAAA 2221  
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2222 GTAAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGTCTGTTTACGAAATAAT 2281  
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 2282 TTGACTCTTCAAAATTTATGATTAACATATGATATCATGCGAAGACGAGAAATTTACTTGG 2341  
 Qy 781 LeuLeuLysGlySerAsnProSerValSerLysLysGluLysIleAsn 796  
 Db 2342 TTGTTAAAGAGATTAATCTTCACTGTAGTAAGGAAAAATAAATAAAC 2389

## RESULT 3

BD063274 2389 bp DNA linear PAT 27-AUG-2002  
 LOCUS Streptococcus pneumoniae antigens and vaccines.  
 BD063274  
 ACCESSION  
 BD063274.1 GI:22608877  
 VERSION JP 2001505415-A/28.  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unclassified.  
 REFERENCE 1 (bases 1 to 2389)  
 AUTHORS Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromocky, J.A.

TITLE  
JOURNAL  
COMMENT

Streptococcus pneumoniae antigens and vaccines  
Patent: JP 2001505415-A 28 24-APR-2001;  
HUMAN GENOME SCIENCES INC  
PN JP 2001505415-A/28  
PD 24-APR-2001  
PF 30-OCT-1997 JP 1998520667  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKY, J PC  
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC  
G01N33/569  
PC G01N33/68  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers.

FEATURES  
source  
1..2389 Location/Qualifiers.  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

ORIGIN  
Alignment Scores:  
Pred. No.: 7,42e-230 Length: 2389  
Score: 4163.00 Matches: 796  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.95% Indels: 0  
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x BD063274 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20  
Db 2 TCTTACGAGTTGGACTGTATCAAGCTAGACGGTTAAGGAAATATTCGGTTTCCTAT 61  
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
Db 62 ATAGATGGAACCAAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGGTTAGCAAG 121  
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60  
Db 122 CGTGAAGGAATCAATGCTGACAAATCGTCATCAAGATAACAGACCAAGGCTATGCTACT 181  
Qy 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80  
Db 182 TCACATGGCGACCACTATCATATTACATGTAAGTTCCCTTATGACGCTATCATCAGT 241  
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
Db 242 GAAGAAATTTACTCATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAAATGAG 301  
Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
Db 302 GTCNAGGGTGGATATGTTATCAAGGTAGATGGAAATACTATGTTTACCTTTAAGGATGCT 361  
Qy 121 AlaHisAlaAspAsnValArgThrLysGluLysIleAsnArgGlnLysGlnGluHisSer 140  
Db 362 GCCCACGCGGATACGTCGTTACAAAGAGAAATCAATCGACAAAACAGACCATAGT 421  
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
Db 422 CAACATCGTGAAGGTGAACCTCCAAAGAAACGATGGTCTGTTCCCTGGCAGCGTTCCGAA 481  
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
Db 482 GGAGCTTACTACAGATGATGGTTATATCTTTAATGCTCTGATATCATAGAGATACT 541  
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
Db 542 GGTGATCTTATATCGTTCTCATGGAGATCATTTACCATTTACATTTCTTAAGAATGAGTTA 601  
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220

602 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCCTATCTGGTCGAGGAAATCTGTCAAAAT 561  
221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
662 TCAGAACCTATCGCGCGCAAAATAGCGATAACATTTCAAGAACAACTGGGTACCTTCT 721  
241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
722 GTAACCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACTAACAGTCAA 781  
261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
782 GCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTGCTTTTGTAGT 841  
281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
842 CAACGACATGTAGAATCTGATGGCCTTCTCTTTGATCGACACAAATCACAACTCGAACA 901  
301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
902 GCTAGAGGTGTTCAGTGCACACGAGATCATTAACCACTTCATCCCTTACTCTCAAAATG 961  
321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
962 TCTGAATTTGGAAGAACGATCGCTGTATATTCCCTTCTGTTATCGTTCAACCATTTGG 1021  
341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
1022 GTACCAGATTCAAGGCCAGAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGGC 1081  
361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
1082 CGCAACCTCGCAACAAATCTTAAATAGACTCAAATTTCTTCTTGGTTAGTCAGCTGTA 1141  
381 ArgLysValGlyGluGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400  
1142 CGAAAGTTGGGAAAGGATATGTAATCGAAGAAAAGGCACTCTCTCGTTATGTTCTTGGC 1201  
401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
1202 AAAGATTATCCATCTGAACTGTTTAAATACTTGAACCAAGTTATCAAACACAGAGAGT 1261  
421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
1262 GTTTTCACACACTTTAACTGCTTAAAGAAAGAAATGCTCTCGTCTGACCAAGAAATTTAT 1321  
441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
1322 GATAAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGNAATTAAGGTCGT 1381  
461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480  
1382 AATTCTGATTTCAAAGCCTTAGACAAATTTATAGAACGCTTGAATGATGAATCGACTAAT 1441  
481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
1442 AAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTAGCACAATTTACCCATCCAGAGCGA 1501  
501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
1502 CTTGGCAACCAAAATCTCAAATTTGATGATGTAAGAACGAGGTTCTGTTATGCTCAATTA 1561  
521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
1562 GCTGATAAGTATACAACGTCAGATGGTTACATTTTGTATGAACATGATATAATCAGTGAT 1621  
541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
1622 GAAGGAGATGCATATGTAAACGCTCATATGGGCGCATAGTCACTGAGTTGAAAAGATAGC 1681  
561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
1682 CTTTCTGATTAAGAAAAGTTGACGCTCAAGCCTATATAAGAAAAGGATATCTTACCT 1741

QY	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr	600	Query Match: 99.95%	Indels: 0		
DB	1742	CCATCTCAGACGCGAGATGTTAAAGCAATCCAACTGGAGATAGTCGACGAGCTATTAC	1801	DB: 6	Gaps: 0		
QY	601	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620	US-09-765-271-56 (1-796) x AX569139 (1-2406)			
DB	1802	AATCGTGTGAAGGGGAAAACGAATTCACCTCGTTCCGACTCCATATATGGTTGAGCAT	1861	QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr	20
QY	621	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640	DB	19	TCTTACGAGTTGGGACTGTATCAAGCTAGAAGCGGTTAAGGAAAATAATCGTGTTCCTAT	78
DB	1862	ACAGTTGAGGTTAAAGCGTAATTTGATTATTCCTCATAGGATCATTAACCAATATT	1921	QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
QY	641	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660	DB	79	ATAGATGGAACACAGCCGACGCAAAACCGAGAAATTTGACTCTCGATGAGGTAGCAAG	138
DB	1922	AAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGAT	1981	QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
QY	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680	DB	139	CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACACAGCAAGGCTATGTCACT	198
DB	1982	TTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAACGTCACATTTCTAATGAT	2041	QY	61	SerHisGlyAspHisTyrHisTyrTrAsnGlyLysValProTyrAspAlaIleIleSer	80
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700	DB	199	TCACATGGCGACCACTATCATTTATTAACATGTAAGGTTCCTTATGAGCTATCATCAGT	258
DB	2042	GGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAAAT	2101	QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
QY	701	LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro	720	DB	259	GAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGAG	318
DB	2102	AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCCGAGAGTCCCT	2161	QY	101	VallLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
QY	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740	DB	319	GTCAGGGTGGATATGTTTATCAAGGTAGATGAAAAATACTATGTTTACCTTTAAGGATGCT	378
DB	2162	CAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAAGCAGAGATTTTGTTCGGA	2221	QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760	DB	379	GCCCAACGGGATAACGTCGTACAAAAGAGGAAATCAATCGACAAAACAAAGACATAGT	438
DB	2222	GTAACGATTTCTAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT	2281	QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780	DB	439	CAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGTGCTTGCCTCGCACGTTTCGCA	498
DB	2282	TTGACTCTTCAAAATATGATAACAAATAGTATCATGGCAGAACGACAAAATTTACTTGG	2341	QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
QY	781	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796	DB	499	GGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT	558
DB	2342	TTGTTAAAGGAAGTAACTCTTCATCTGTAAGTAAAGAAAAATAAAC	2389	QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
RESULT 4	AX569139	Sequence 2347 from Patent WO02077021.	PAT 29-NOV-2002	DB	559	GGTGATGCTTATATCGTCTCATGGAGATCATTAACATACATTCCTAAGAAATAGTTA	618
DEFINITION	AX569139	2406 bp DNA linear		QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
ACCESSION	AX569139.1	GI:26002636		DB	619	TCAGCTAGCGAGTTGGCTGCTGCAGAGGCTTCCTATCTGGTCGAGGAAATCTGTCAAAT	678
VERSION	KEYWORDS	Streptococcus pneumoniae		QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
SOURCE	ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		DB	679	TCAAGAACCTATCGCCGACAAAATAGCGATTAACACTTCAAGAACAACTGGGTACCTTCT	738
REFERENCE	AUTHORS	Magisani, V., Tettelin, H. and Fraser, C.		QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
TITLE	JOURNAL	Streptococcus pneumoniae proteins and nucleic acids		DB	739	GTAAGCAATCCAGGAATCAAAATACTAACAAGCAACAAACAGCAACACTACAGTCAA	798
FEATURES	source	Location/Qualifiers		QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
		1. 2406		DB	799	GCAAGTCAAGTAATGACATTGATGATCTCTTGAAACAGCTCTACAACTGCCTTTGAGT	858
		/organism="Streptococcus pneumoniae"		QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
		/mol_type="unassigned DNA"		DB	859	CAACGACATGTAGAAATCTGATGGCCTTGTCTTTGATCCAGCAACAAATCAAACTCGAACA	918
		/db_xref="taxon:1313"		QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
Alignment Scores:				DB	919	GCTAGAGGTTCGAGTGCACACGGAGATCATTAACATTCATCCCTTACTCTCAATG	978
Pred. No.:	7.48e-230	Length:	2406	QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
Score:	4163.00	Matches:	795	DB	979	TCTGAAATTGAAGAACGAATCGTCTGATATTATCCCTTCGTTATCGTTCAACACCATGG	1038
Percent Similarity:	99.87%	Conservative:	0				
Best Local Similarity:	99.87%	Mismatches:	1				

QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 DB 1039 GTACCAAGTTCAGGCCAGAACCAACCAAGTCCACAAACGACCTCGGACCTAGTCCAGGC 1098  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerLeuValSerGlnLeuVal 380  
 DB 1099 CCGCAACCTCGACCAAACTCTAAATAGACTCAAAATCTCTCTTGGTTAGTCAGCTGGTA 1158  
 QY 381 ArgLysValGluGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 DB 1159 CGAAAGTTCGGGAAGGATATGATTTCGAAGAAAGGGCATCTCTCTGTTATGCTTTGGC 1218  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuGluSer 420  
 DB 1219 AAGATTTACATCTGAACTGTAAATCTTGAAGCAAGTTATCAAAACAGAGAGT 1278  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 DB 1279 GTTTTCACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTAT 1338  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 DB 1339 GATAAGCATATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTAATAAAGGTCGT 1398  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480  
 DB 1399 AATCTGATTTCCAGCTTAGACAAATATTAGAACGCTTGAATGATGATCACTAAT 1458  
 QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProLeuThrHisProGluArg 500  
 DB 1459 AAAGAAAAATGTTAGATGATTTATTGGCATTTCTAGCACCAATATCCCATCCAGAGCGA 1518  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrLysAspGluValArgIleAlaGlnLeu 520  
 DB 1519 CTTGGCAACCAAAATCTCAAAATGAGTACTAGAACGAGGTCGTATGCTCAATTA 1578  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 DB 1579 GCTGATAAGTATACACGTCAGATGGTTACATTTTGTGATGAACATGATATATCAGTGAT 1638  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 DB 1639 GAAGGAGATGCATATGTAACGCTCATATGGCCATAGTCACTGGATTGGAAAGATAGC 1698  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 DB 1699 CTTTCTGATAAGGAAAAAGTTGACGCTCAAGCCTATATAAGAAAAAGGATATCTCACT 1758  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 DB 1759 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCTATTATC 1818  
 QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 DB 1819 AATCGTGTGAAGGGGAAAAACCAATTCCTCACTCGTTCGACTTCATATATGGTTGAGCAT 1878  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640  
 DB 1879 ACAGTTGAGGTTAAAAACCGGTAAATTTGATTATTTCTCATAGGATCATTAACCAATAAT 1938  
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 DB 1939 AAATTTGCTGGTTGATGATACACATACAAAGCTCCAAATGGCTATATCTTTGGAGAT 1998  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 DB 1999 TTGTTTGGCAGGATTAAGTACTAGTAGAACACCCCTGACGACCGTCCACATTTCTAATGAT 2058  
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 DB 2059 GGAATGGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2118  
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720

DB 2119 AAGAACTTCAAGCGGATGAGAGCCAGTAGAGAAACACACTGCTGAGCCAGAAAGTCCCT 2178  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740  
 DB 2179 CAAGTAGAGACTGAAAAAGTAGAAGCCAACTCAAGAGCAGAAAGTTTGTCTTGCAGAA 2238  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 DB 2239 GTAACGGATTCTAGTCTGAAAGGCCAATGCAACAGAAACTCTAGCTGTTTACGAAATAAT 2298  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780  
 DB 2299 TTGACITCTTCAATATTGATTAACAATAGTATCATGCGAAGCAGAAAAATTTACTTCG 2358  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 DB 2359 TTGTTAAAGGAAGTAACTCTTCACTGTTAAGTAAGGAAAAATAAATAC 2406  
 RESULT 5  
 BD268048  
 LOCUS Streptococcus pneumoniae protein and immunogenic fragments for  
 DEFINITION vaccines.  
 ACCESSION BD268048  
 VERSION BD268048.1 GI:33077816  
 KEYWORDS JP 2002532561-A/6.  
 SOURCE Streptococcus pneumoniae  
 ORGANISM Streptococcus pneumoniae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 2451)  
 AUTHORS Johnson,L.S., Koenig,S. and Adamou,J.E.  
 TITLE Streptococcus pneumoniae protein and immunogenic fragments for  
 JOURNAL Patent: JP 2002532561-A 6 02-OCT-2002;  
 MEDIMMUNE INC  
 COMMENT OS Streptococcus pneumoniae  
 PN JP 2002532561-A/6  
 PD 02-OCT-2002  
 PF 21-DEC-1999 JP 2000589215  
 PR 21-DEC-1998 US 60/113048  
 PI LESLIE S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU  
 PC A61K39/09,A61K38/00,A61P31/10,A61K37/02  
 CC Streptococcus pneumoniae protein and immunogenic fragments for  
 CC vaccines  
 FH Key Location/Qualifiers  
 FT source 1..2451  
 FT /organism='Streptococcus pneumoniae'.  
 FEATURES  
 source Location/Qualifiers  
 1..2451  
 /organism="Streptococcus pneumoniae"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1313"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,66e-230 Length: 2451  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-765-271-56 (1-796) x BD268048 (1-2451)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 DB 61 TCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAAATATCGTGTTCCTAT 120  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 DB 121 ATAGATGGAAACAACGACGCGCAAAAAACGGAGAAATTTGACTCTGATGAGGTTAGCAAG 180



QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleIleThrAspGlnGlyTyrValThr 60  
 Db 181 CGTGAAGAAATCAATCTGAGCAATCTCATCAGATTAACAGACCAAGGCTATGCTACT 240  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 241 TCACATGCGCACCATATCATTTATTAATGTAAGGTTCCTTATGACGCTATCATCAGT 300  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 301 GAAGAATTTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGGATTTGTTAATGAG 360  
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120  
 Db 361 GTCAAGGTTGATGATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGATGCT 420  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db 421 GCCCACGCGGATAACGTCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 480  
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 481 CAACATCGTGAAGGTGGAACTCCAGAAACGATGTGTGCTTGTGCTTGGCACGTTCCGCAA 540  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 541 GGACGCTATACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATACT 600  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 601 GGTGATGCTTATATGCTTCTCATGGAGATCATTTACCATTAATCTTCTTAAGAATGAGTTA 660  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 661 TCAGCTAGCGAGTTGCTGCTCAGAAAGCTTCTTATCTGTCGAGGAAATCTGCAAAAT 720  
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 Db 721 TCAGAACCTATCGCGACAAATAGCGATAACACTTCAGAACCAACCTGGGTACTTCT 780  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 781 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACCAACAGCAACACTAACAGTCAA 840  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 841 GCAAGTCAAGTAATGACATTTGATGCTCTTGAACACAGCTCTACAACTGCTCTTGAGT 900  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 901 CAACGACATGTAGAAATCTGATGCTTGTCTTTGATCCAGCACAAATCACAGTCCGAACA 960  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 961 GCTAGAGGTGTGCGATGTCACACGAGATCATTTACCACTTCATCCCTTACTCTCAAAATG 1020  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrP 340  
 Db 1021 TCTGAATTTGAAGAAGAAATCGCTCGTATTTATTCCTTCTGTTATCTGTTTCTTGAACCATTTGG 1080  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1081 GTACACGATTTCAAGGCCAGAACCAACCAAGTCCACAAAGGACTCCGGACCTAGTCCAGGC 1140  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1141 CGCAACCTGCAACCAATCTTAAATPAGACTCAAAATCTCTTCTTGTGTAGTCAGCTGCTGA 1200  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1201 CGAAAAAGTTGGGAAGGATATGTTTTCGAAGAAGGGCATCTCTCGTTATGCTTTGGC 1260

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 1261 AAAGATTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 1320  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1321 GTTTCACACACTTAACTGCTAAAAAGAAAAATGTTGCTCTCGTGCACCAAGAAATTTAT 1380  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1381 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGGCTTGTGTGNAATAAAGGTCGT 1440  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1441 AATTCTGATTTCCAAAGCCTTAGACAAATTTATTAGAAGCCTTGAATGATGAATCGACTAAT 1500  
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1501 AAAGAAAAATTTGGTAGATGATTTATTGGCATTTCTTAGCAACCAATTTACCATCCAGAGCGA 1560  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1561 CTTGCAAAACCAATTTCTCAAAATGAGTATACTGAAGACGAGTTCTGATTTGCTCAATTA 1620  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1621 GCTGATAAGTATACAACTGATGTTTACATTTTTTGTATGAACATGATATATATCAGTGAT 1680  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560  
 Db 1681 GAAGAGATGATGATGTTAAACGCTCATATGGCCCATGTCATCTGATTTGGAAGATAGC 1740  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1741 CTTTCTGATAGAAAAAGTTGCGACTCAAGCTATATCTAAAGAAAAAGGTATCTTACT 1800  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1801 CCATCTCAGACGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGTATTTAC 1860  
 QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1861 AATCGTGTGAAGGGGAAAAACGAATTCACCTGCTTCCACTTCCATATATGTTGAGCAT 1920  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1921 ACAGTTGAGTTAAAAACGGTAATTTGATTTATCTCTCATAGGATCATTCATTAATAT 1980  
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1981 AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT 2040  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 2041 TTGTTTGGCAGATTAAGTACTACGTAGAACACCCCTGACGACGTCCTCAATTTCTAATGAT 2100  
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2101 GGATGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGACCAACAGATGAGATCCAAAT 2160  
 QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2161 AAGAACTTCAAAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCAGAGTCCCT 2220  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2221 CAAGTAGAGACTGAAAAAGTAGAACCCCACTCAAAAGAAAGCAAGAGTTTGTCTTGGAAA 2280  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2281 GTACCGATTTCTAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGTTGTTTACGAAATAAT 2340  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780



```

Db      2341 TTGACTCTTCAAAATTATGGATAACATAGTATCATGGCAGGACGAGAAAATTACTTGG 2400
Qy      781  LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db      2401 TTGTTAAAGGAAGTAATCCCTTCACTCTGTAAGTAAGGAAAAATAAAC 2448

RESULT 6
LOCUS      AR344447          2451 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 9 from patent US 6582706.
ACCESSION AR344447
VERSION   AR344447.1  GI:33740486
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2451)
AUTHORS     Johnson, L.S. and Adamou, J.E.
TITLE       Vaccine compositions comprising Streptococcus pneumoniae
            polypeptides having selected structural MOTIFS
JOURNAL     Patent: US 6582706-A 9 24-JUN-2003;
FEATURES    Location/Qualifiers
            source
            1..2451
            /organism="unknown"
            /mol_type="genomic DNA"

ORIGIN
Alignment Scores:  7.66e-230      Length: 2451
Pred. No.: 4163.00      Matches: 796
Score: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 99.95%      Indels: 0
DB: 6      Gaps: 0

US-09-765-271-56 (1-796) x AR344447 (1-2451)

Qy      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db      61  TCTTACGAGTTGGACTGTATCAAGCTAGACGGTTAAGGAAAATAATCGTGTTTCCTAT 120
Qy      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      121 ATAGATGGAACAAACGAGCAGCAAAAACGAGAAATTGACTCTCTGATGAGTTAGCAG 180
Qy      41  ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      181 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAGGCTATGTCACT 240
Qy      61  SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      241 TCACATGGCGACCATATCATATTACATGGTAAAGGTTCTTATGACGCTATCATCAGT 300
Qy      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db      301 GAAGAATAACTCATGAAGAAGTCCAAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG 360
Qy      101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      361 GTCNAGGGTGATATGTTATCAAGGTAGATGGAAAATACTATGTTTACCTTAAGGATGCT 420
Qy      121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db      421 GCCCAGCGGATAACGTCCTGCTACAAAGAGGAATCAATCGCAAAAACAGAGCATAGT 480
Qy      141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db      481 CAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTGCTGTGCTTGGCCATCGCTCGCAA 540
Qy      161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      541 GGACGCTTACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 600

```

```

Qy      181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      601 CGTGATGCTTATATATGTTCTCATGGATCATTAACCATTAATCTAAGAATGAGTTA 660
Qy      201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      661 TCAGCTAGCGAGTTGGCTGCTCAGAAGCTTCTCTATCTGGTCGAGGAATCTGTCAAT 720
Qy      221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      721 TCAGAACCTATCGCCGCAAAATAGCATACACTTCAAGAACAACTGGGTACTCTTCT 780
Qy      241 ValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      781 GTAGCAATCCAGGAACCTACAAATATAACAACAGCAACAAACAGCAACACTAACAGTCAA 840
Qy      261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProIleSer 280
Db      841 GCAAGTCAAGTATATGACATTTGATAGTCTCTTGAACACAGCTCTACAACTGCCTTTGAGT 900
Qy      281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      901 CAACGACATGTAGAAATCTGATGGCTTCTTTGATCCAGCACAAATCACAGTCCGAACA 960
Qy      301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      961 GCTAGAGGTGTTCAGTGCACACGAGATCATTAACCATTTCACTCCCTTACTCTCAAAATG 1020
Qy      321 SerGluLeuGluGlyArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db      1021 TCTGATTTGGAGAACGAAATCGCTGATATATCCCTTCTGTTATCGTTCAAACCATTTGG 1080
Qy      341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      1081 GTACGAGATTCAAGGCCAGAACCAACCAAGTCCACACGACTCCGACACTAGTCCAGGC 1140
Qy      361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db      1141 CCGCAACCTCGACCAAAATCTTAAATAGACTCAAAATCTCTCTTTGGTTAGTCAGCTGTA 1200
Qy      381 ArgLysValGlyGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400
Db      1201 CGAAAAGTTGGGAGGAGATATGATATTCGAAGAAAAGGCGATCTCTCGTTATGCTTTGGC 1260
Qy      401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db      1261 AAGAAATTACCATCTGAAACTGTTTAAATACTTTGAAAGCAAGTTATCAAAACAGAGAGT 1320
Qy      421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db      1321 GTTTCACACACTTTAACTGCTTAAAGAAAAGAAAATGTTGCTCTCTGACCAAGAAATTTAT 1380
Qy      441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db      1381 GATAAGCATATATATCTGTTAACTGAGGCTCAATAAGCCCTTGTTTGAAAATAAGGGTCGT 1440
Qy      461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db      1441 AATTCTGATTCCAGCCCTTAGACAAATATTATAGACGCTTGAATGATGATCGACTAAT 1500
Qy      481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db      1501 AAAGAAAAATTGGTAGATGATTTTATTTGGCATCTCTAGCACCAATTTACCCATCCAGAGCGA 1560
Qy      501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db      1561 CTTGGCAAAACCAAAATCTCAAATTTGAGTATCTGGAAGACGAAAGTTCGTATTTGCTCAATTA 1620
Qy      521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db      1621 GCTGATAAGTATACACAGCTCAGATGGTTACATTTTTTGATGAAATGATATATATCATGAT 1680

```

Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer 560  
 Db 1681 GAAGGAGATGCATATGTAACCCCTCATATGGCCCATAGTCTAGTGGTAAAGATAGC 1740  
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLysLeuPro 580  
 Db 1741 CTTTCTGATAGGAAAGAGTTGCAGCTCAAGCTATCTAAGAAAAAGGTATCTACCT 1800  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1801 CCATCTCCAGACCGCAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTAC 1860  
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1861 AATCTGTGTAAGAGGGAAGAAAGAAATCCACCTGCTGCTGCTTCCATATATGTTGAGAT 1920  
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1921 ACAGTTGAGTTAAACCGTAAATTTGATTTATCTCATAGGATCATTAACATAATATT 1980  
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1981 AAATTTGCTTGTGTTGATCATCACATACAAAGCTCCAAATGGCTATACCTTGGAGAT 2040  
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 2041 TGTGTTGCCAGCATTAAGTACTACGTAGTAAACCCCTGACGACGCTCCACATCTAATGAT 2100  
 Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2101 GGATGGGGCAATCCCACTGAGCATGTGTTAGGCAAGAAAGACCAAGTGAAGTCAAT 2160  
 Qy 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2161 AAGAACTTCAAGCGCATGAAGCAGCAGTACGAGTAAACACCTGCTGAGCAGAGTCCCT 2220  
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740  
 Db 2221 CAAGTAGAGCTGAAAGAGTAGAAGCCCACTCAAGAGAGCAGAGTTTGTGTCGAA 2280  
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2281 GTAACGGATCTTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGTTTACGAAATAAT 2340  
 Qy 761 LeuThrLeuGlnIleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 2341 TTGACTCTTCAAAATTTATGATAACAATAGTATCATGCGAAGCAGAGAAATTACTTGG 2400  
 Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 2401 TTGTTAAAGGAAGTAACTCTTCTATCTGTAAGTAAGGAAATAAATCAAC 2448  
 RESULT 7  
 AF291695 2541 bp DNA linear BCT 15-MAR-2001  
 LOCUS Streptococcus pneumoniae pneumococcal histidine triad A protein  
 DEFINITION (phtA) gene, complete cds.  
 ACCESSION AF291695 GI:13345012  
 VERSION AF291695.1  
 KEYWORDS  
 SOURCE Streptococcus pneumoniae  
 ORGANISM Streptococcus pneumoniae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 1 (bases 1 to 2541)  
 Wizenmann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C.,  
 Gayle,A., Barash,S.C., Rosen,C.A., Masure,H.R., Tuomanen,E.,  
 Gayle,A., Brewah,Y.A., Walsh,W., Barren,P., Lathigra,R., Hanson,M.,  
 Langermann,S., Johnson,S. and Koenig,S.  
 Use of a whole genome approach to identify vaccine molecules  
 affording protection against Streptococcus pneumoniae infection  
 Infect. Immun. 69 (3), 1593-1598 (2001)  
 MEDLINE 21116976

11179332  
 2 (bases 1 to 2541)  
 CHOI,G.H.  
 Direct Submission  
 Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,  
 Inc., 9410 Key West Ave., Rockville, MD 20850, USA  
 Location/Qualifiers  
 1..2541  
 /organism="Streptococcus pneumoniae"  
 /mol\_type="genomic DNA"  
 /strain="N4"  
 /db\_xref="taxon:1313"  
 91..2541  
 /gene="phtA"  
 91..2541  
 /gene="phtA"  
 /note="phtA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="pneumococcal histidine triad A protein"  
 /protein\_id="AAK19155.1"  
 /db\_xref="GI:13345013"  
 /translation="MKINKKYLVSAAALILSVCSYELGLYQARTVKNRVSVIDGK  
 QATKTEMLTPEVSKREGINAEQIVIKITDQYVTSFGDHYVYNGKVPYDAIIEE  
 LLMKDPNKLKDEDIVNEVGKIVIKDKIYVYLKDAHADNVTKELINRQKQHS  
 QHREGTTPNDGAVALARSGRYTDDGVIENASDIIEBTGDAYIVPHGDHHTYIPKN  
 ELSASELAFAFLSGRNLNSRTYRQNSDGLVDFPAQITSRTAGVAPRHGDHYHF  
 TNSOASQNDISLLQLYKLPISORHVESDGLVDFPAQITSRTAGVAPRHGDHYHF  
 TPYSSQLEERIEARIILPVRSHNVDPDSRPEQSPPTPEPSPQAPAPKIDSN  
 SSLVQLVRKGVGEGYVPEKGLSRVYFAKDLPSSETVKNLESKLSKQESVHTLTAKKE  
 NVAPDQBFYDKAYNLLFEAKHAFNKGNSDFQALDKLLELNDESNKELVDDL  
 LAFLAPITHEPLRNPQIEYTEDVEVRIADKYTSDGYIFDEHDIISDEGDYV  
 TFMGHSHWIGKDSLKDEKVAQAQYTKKGLPPSPDADVKANPTGDSAAIYNRVK  
 GFKPIVLRLPYMVEHTVKNLIIIPKDHVHNIKFAWFDHHTYKAPNGTLEDF  
 ATIKYVEHPDERPHSDNGWASRHLGKDKHSDPNKNFKADEEPVETPAEPEVP  
 QVETEKVQLKEAEVLAKVYDSSILKANATHTLAGLRNLLTIQIMDNNSIWAEEKL  
 LALLKGNPSVSVSEKIN"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8,01e-230 Length: 2541  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-765-271-56 (1-796) x AF291695 (1-2541)  
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 151 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTAT 210  
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 211 ATAGATGGAAAAACAAGCAGCGCAAAACGGAAATTTGACTCCTCTGATGAGGTAGCAAG 270  
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 271 CGTGAAGGATCAATCGCTGAGCAATCGTCATCAAGATAACAGCAAGGCTATGTCAC 330  
 Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 331 TCACATGGCGACCATCATATTATTATGTAAGGTTTCCTTATGACGCTATCATCAGT 390  
 Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 391 GAAGAAATCTCATGAAGATCCAACTAATGCTAAAGATGAGGATATTGTTAATGAG 450  
 Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 451 GTCAGGCTGATATGTTATCAAGGTAGTAGGAAATACTATGTTTACCTTAAAGGATGCT 510

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGlnHisSer 140  
DB 511 GCCACCGCGATAACGTCGTAACAAAGAGGAAATCAATCGACAAACAAAGAGCATAGT 570  
QY 141 GlnHisArgGluGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
DB 571 CAACATCGTGAAGGTGGAATCTCAAGAAACGATGGTGTCTGCTTGGCCTTGCACGTTCCGCAA 630  
QY 161 GlyArgThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
DB 631 CGACGCTATACAGATGATGGTATATCTTTAATGCTTCTGATATCATAGAGATACT 690  
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
DB 691 GGTGATGCTTATATCGTCTCATCGAGATCAATACCATTAATCTTCAAGATCAGTTA 750  
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
DB 751 TCAGCTAGCGAGTTGGCTGCTGCGAAGCCTTCTTATCTGGTTCGAGGAAATCTGTCAAAAT 810  
QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
DB 811 TCAGAACCTATCGCCGACAAATAGCGATTAACACTTCAAGAACAAACTGGGTACCTTCT 870  
QY 241 ValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
DB 871 GTAAAGCAATCCAGGAATCAAAATACTTAACACAGCAACACAGCAACTTAACAGTCAA 930  
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
DB 931 GCAAGTCAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 990  
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
DB 991 CAACGACATGTAGATCTGATGGCTGTCTTGTGATCAGCACAAATCACAAAGTCGAACA 1050  
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
DB 1051 GCTAGAGGTGTGAGTGGCAGTCCACAGGATCATTTACCACTTCACTCCCTTACTCTCAATG 1110  
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
DB 1111 TCTGAATGGAAGAACCAATCGCTGCTGATTTATTCCTCTGCTTATCTGTTCAACCAATGG 1170  
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGlnProSerProGly 360  
DB 1171 GTACCAATTCAGCCAGACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1230  
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
DB 1231 CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTCTTTGTTAGTCAGCTGTTA 1290  
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
DB 1291 CGAAAGTTGGGAAGATATGATTTGGAAGAAAGGGCATCTCTGCTTATGCTTTGCG 1350  
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnGluSer 420  
DB 1351 AAAGATTTACCACTGAACTGTAAATCTTGAAGAGCAAGTTTCAAAACCAAGAGAT 1410  
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
DB 1411 GTTTCACACTTAACTGCTAAAGAAAGAAATGTTGCTCCTCGTGACCAAGAAATTTAT 1470  
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
DB 1471 GATAAGCATATATCTGTTAATCGAGGCTCAATAAGCCCTTGTGTGAAATAAGGTCGT 1530  
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
DB 1531 AATTCTGATTTCAAGCCTTAGCAAAATATTAGAACGCTTGAATGATGATGATGATGATGAT 1590  
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500

DB 1591 AAAAGAAAATTTGGTAGATGATTTATTTGGCATTTCTTAGCACCAATTAACCATCCAGACGA 1650  
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
DB 1651 CTTGGCAACCAAAATCTCAAAATGAGTATCTGAAGACGAAGTTCTGATTTCTCAATTA 1710  
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSerAsp 540  
DB 1711 GCTGATAAGTATACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1770  
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560  
DB 1771 GAAGAGATGATATATGTAACGCTCATATGGCCATAGTCTCATGATTTGGAAGATAGC 1830  
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
DB 1831 CTTTCTGATAAGGAAAGTTGACGCTCAGCCTATATCTAAAGAAAAGGATTCCTACCT 1890  
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
DB 1891 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1950  
QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
DB 1951 AATCGTGTGAAGGGGAAACGAAATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2010  
QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
DB 2011 ACAGTTCAGGTTAAACCGTAAATTTGATTTATCTCTATAAGGATCATTACCATTAAT 2070  
QY 641 LysPheAlaThrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
DB 2071 AAATTTCTGTTTGTATGATCATACATACAAAGCTCCAAATGCTATACCTTGGAGAT 2130  
QY 661 LeuPheAlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
DB 2131 TTGTTTCGAGCATTAAGTACTACGTAGAACACCCCTCCGAAACCTCCACATTTCTAATGAT 2190  
QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
DB 2191 GGATGGGCAATGCCAGTGGAGCATGTTTAGGCAAGAAAGACCACTGAAGATCCAAAT 2250  
QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
DB 2251 AAGAATCTCAAGCGGATGAAGCCAGTGAAGGAAACACCTGCTGAGCCAGAGTCCCT 2310  
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
DB 2311 CAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTGTCTGCGAAA 2370  
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
DB 2371 GTAACGATTTCTAGTCTGAAAGCCATGCAACAGAACTCTAGCTGTTTACCAATPAT 2430  
QY 761 LeuThrIleGlnIleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
DB 2431 TTCACTCTTCAATTTAGTAACTAGTATCATGCGAAGCAGAAAAATTAATTCTTGC 2490  
QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
DB 2491 TTGTTAAAGGAAGTAATCCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2538

## RESULT 8

AR218862

LOCUS

DEFINITION

SEQUENCE

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

8195 bp DNA linear PAT 25-SEP-2002

Sequence 94 from patent US 6420135.

AR218862

AR218862.1 GI:23319796

Unknown.

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 8195)  
 AUTHORS Kunsch, C.A., Choi, G.H., Dillion, P.S., Rosen, C.A., Barash, S.C.,  
 Fannon, M.R. and Dougherty, B.A.  
 TITLE Streptococcus pneumoniae polynucleotides and sequences  
 JOURNAL Patent: US 6420135-A 94 16-JUL-2002;  
 FEATURES Location/Qualifiers  
 source 1..8195  
 /organism="unknown"  
 /mol\_type="genomic DNA"

ORIGIN

Alignment Scores: 3,48e-229 Length: 8195  
 Score: 4163.00 Matches: 795  
 Percent Similarity: 99.87% Conservative: 0  
 Best Local Similarity: 99.87% Mismatches: 1  
 Query Match: 99.95% Indels: 0  
 DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AR218862 (1-8195)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20  
 Db 3054 TCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTAT 3113

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 3114 ATAGATGGAAACAACGACGACGCAAAACGGAAGATTTGACTCCCTGATGAGTTAGCAAG 3173

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 3174 CQTGAAGGATCAATGCTGAGCAATCGTCATCAAGATTAACAGACCAAGGCTATGTCACT 3233

QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleLeuSer 80  
 Db 3234 TCACATGGCGACCATCATCATTTACAAATGATGGTAAAGTTTCCTTATGACCTATCATCAGT 3293

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 3294 GAAGAAATTAATCATGAAGATCCAACTATAAGTAAAGATGAGGATATGTTAATGAG 3353

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAspAla 120  
 Db 3354 GTCAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGCT 3413

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140  
 Db 3414 GCCCAGCGGATAACCTCCGTACAAAGAGAGAAATCAATCGCAAAACAAGAGCATAGT 3473

QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyValAlaValAlaAlaArgSerGln 160  
 Db 3474 CAACATCGTGAAGGTGGAACTCCAGAAGACGATGGTGTGCTTGGCCTGGCAGCTTCGCA 3533

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuAspThr 180  
 Db 3534 GGAAGCTATACATACATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATAC 3593

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 3594 GGTGATGCTTATATCTCTCTCATGGAGATCATTACCATTTACATTCCTAAGAATGAGTTA 3653

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 3654 TCAGCTAGCGAGTTGGCTGCTGCAAGAGCCTTCTCTATCTGCTGAGGAATCTGTCAAT 3713

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 3714 TCAGAACCTTATCGCGACAAAATAGCGATAACACTTCAAGAAACAAACTGGGTACCTTCT 3773

QY 241 ValSerAsnProGlyThrAsnThrAsnThrAsnThrSerAsnAsnSerAsnSerGln 260  
 Db 3774 GTAAGCAATCCAGGAATCAAAATACTAAACCAAGCAACCAACCAATCAACAGTCAA 3833

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 3834 GCAAGTCAAGTAAATGACATTTAGTAGTCTCTTGAACAGCTCTACAACTCCCTTTGAGT 3893

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 3894 CAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCACAAAGTCGAACA 3953

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 3954 GCTAGAGTGTTCAGTGCCACACGGAGATCATTACCATTCACTCCCTTACTCTCAAAATG 4013

QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 4014 TCTGAATTTGGAAGAACGAATCGCTCGTATTATTCCTTCGTTCGTTCCTTCAACCATTTGG 4073

QY 341 ValProAspSerArgProGluGluProSerProGlnProThrProGluProSerProGly 360  
 Db 4074 GTACCAGATTCAAGGCCAGAACCAAGTCCACACCCAGCTCCGGAACCTAGTCCAGGC 4133

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerSerAsnSerLeuValSerGlnLeuVal 380  
 Db 4134 CGGCAACCTGACCAAAATCTTAAATACACTCAAAATCTCTCTTGTGTAGTCAGCTGTA 4193

QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 4194 CGAAAGTTGGGGAAGGATATGATTCGAAGAAAGGGCATCTCTCGTTATGCTCTTTGGC 4253

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 4254 AAAGATTTACCATTCTGAACTGTGTTAAATCTTTGAAGCAAGTTATCAAAACAAGAGAGT 4313

QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 4314 GTTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCTCTCTGAGCAACAAATTTAT 4373

QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 4374 GATAAAGCATATAATCTGTAACTGAGGCTCAATAAGGCTTGTGTGAAATTAAGGTCGT 4433

QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 4434 AATCTGATTTCCAGCCTTAGACAAATTTATTAAGACGCTTGAATGATGATCGACTAAT 4493

QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 4494 AAAGAAATTTGTAGATGATTTATTTGCAATTCCTAGCACTTACCCATCCAGAGCGA 4553

QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 4554 CTTGGCAACCAAAATCTCAAAATTTGAGTATCTGAGAGCAAGTTCGTATTGCTCAATTA 4613

QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540  
 Db 4614 GCTGATAAGTATACACGTCAGATGTTTACATTTTGTGACATCATATATAATCAGTGAT 4673

QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 4674 GAAGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTTGGAAAGATAGC 4733

QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 4734 CTTTCTGATAAGGAAAGTTGAGCTCAAGCCTATACAAAGCAAGATATCTTACCT 4793

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 4794 CCATCTCCAGACCGCATGTTAAAGCAATCCAACTGGAGATAGTCAGCAGCATATTATC 4853

QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 4854 AATCGTGTGAAGGGGAAACCAAAATTCACCTCGTTCGATTCATATATATGTTGAGCAT 4913

QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640

```

Db 4914 ACAGTTGAGGTTAAACCGGTAAATTTGATTATCTCTAAGGATCATACCAATAAT 4973
QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProLenGlyTyrThrLeuGluAsp 660
Db 4974 AAATTTCTTGGTTTGAATGATCACATACAAAGCTCAAATGGCTATACCTTGAAGAT 5033
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 5034 TTGTTTCGCGAGATTAACTACTAGTAAACACCCCTGACGAACCTCCACATTTCTAATGAT 5093
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 5094 GGATGGCGCAATGCCAGTGCAGCTGTGTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 5153
QY 701 LysAsnPhelLysAlaAspGluGluProValGluThrProAlaGluProGluValPro 720
Db 5154 AAGAACCTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 5213
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740
Db 5214 CAAGTAGAGCTGAAAAGTAGAGGACCACTCAAGAGCAGAGATTTTGTCTTGCGAAA 5273
QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 5274 GTAAACGATTCTAGTCTGAAAGCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 5333
QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780
Db 5334 TTGACTCTTCAAAATTTATGGATAACAATAGTAGTATCATGGCAGAGCAGAAAAATTTACTTGC 5393
QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleLeu 796
Db 5394 TTGTTAAAGAAAGTAACTCTTCACTCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 9
BD003774
LOCUS 8195 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003774
VERSION BD003774.1 GI:18631735
KEYWORDS JP 2001501833-A/94.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C., Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;
COMMENT HUMAN GENOME SCIENCES INC
OS Unidentified
PN JP 2001501833-A/94
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH.
P1 MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19, C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..8195 /organism='Unidentified'.
FT Location/Qualifiers
1..8195
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
FEATURES
Source

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 3,48e-229 Length: 8195
Score: 4163.00 Matches: 795
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x BD003774 (1-8195)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 3054 TCTTACAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAAATAATCGTGTTCCTAT 3113
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 3114 ATAGATGGAAAAACAAGCGACGCAAAAACCGAAGATTTTGACTCTCTGATGAGGTAGCAAG 3173
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 3174 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCCCAAGCTATGCTACT 3233
QY 61 SerHisGlyAspHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 3234 TCACATGGCGACCACTATCATATTATTAATGTTAAGGTTCTCTTATGACGCTATCATCAGT 3293
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 3294 GAAGAATTACTCATGAAGATCCAACTATPAAGCTAAAGATGAGGATATTTGTAATGAG 3353
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 3354 GTCAAGGTTGGATATGTTATCAAGGTAGATGGAATACTACTATGTTTACCTTAAGGATGCT 3413
QY 121 AlaHisAlaAspAsnValArgThrLysGluLysIleAsnArgGlnLysGlnLysHisSer 140
Db 3414 GCCCAGCGGATTAACCTCGGTACAAAAGAGGAATCAATCGACAAAACAAGAGCATAGT 3473
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db 3474 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTCTGCTTGGCCTGGCACGTTCCGAA 3533
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 3534 GGACGCTATACATACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 3593
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 3594 GGTGATGCTTATATCTGTTCTCATGGAGATCATTAACCATTTACATTCTTAAAGATGAGTTA 3653
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyValAsnLeuSerAsn 220
Db 3654 TCAGCTAGCGAGTTGGCTGCTCGAGNAGCTTCTCTATCTGCTGAGGAAATCTGTCAAAT 3713
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 3714 TCAAGAACCTATCGCGCAGCAAAATAGCGATAACACTTCAAGAACCAAACTGGGTACCTTCT 3773
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 3774 GTAAGCAATCCAGGAACATAAATACTAACACAAGCAACACAGCAACACTAACAGTCAA 3833
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 3834 GCAAGTCAAGTAATGACATTTGATGATCTCTTTGAAACAGCTCTACAACTGCTTTGAGT 3893
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 3894 CAACGACATGTAGAATCTGATGGCCTTGTCTTTTGTATCCAGCACAAATCACAGTCAACA 3953
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320

```

Db 3954 GCTAGAGTGTGAGTCCACGAGGATCATTTACCATTCCTTACTCTCAATG 4013  
 Qy 321 SerGluLeuGluArgIleAlaAArgIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 4014 TCTGAATTGGAAGAACGAATCGCTCGTATTATTTCCCTTCGTTATCGTTCAACCATTTGG 4073  
 Qy 341 ValProAspSerArgProGluInProSerProGlnProThrProGluProSerProGly 360  
 Db 4074 GTACCAAGATTCAAGCCAGAACCAACAGTCCACCGACTCCGGAACTTAGTCCAGGC 4133  
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 4134 CCGCAACTGCACCAATCTTAATAAGACTCAATCTCTTTGGTTAGTCAGTCGGTA 4193  
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAla 400  
 Db 4194 CGAAAGTTGGGAAGGATATGTATTCAAGAAAGGGCATCTCTCGTTATGTCTTTGGC 4253  
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnLys 420  
 Db 4254 AAGATTATACCATCTGAACACTGTAAATACTTTGAAGCAAGTTATCAAAACAAGAGAGT 4313  
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 4314 GTTTCACACACTTAACCTGCTAAAGAAAGAAATGTGTCTCTCGTGACCAAGAAATTTAT 4373  
 Qy 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 4374 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTTGTTGAATAATAAGGGTCGT 4433  
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 4434 AATTCGTGATTTCCAGCCCTTAGACAAATTTATAGAACCTTGATGATGATGATGATGAT 4493  
 Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 4494 AAGAAATAATGGTAGATGATTTATGGCATTCCTAGCACCAATTTACCATCCAGAGCGA 4553  
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 4554 CTTGGCAACCAACCAATTTCTCAATTTAGTATATCTAGAACGAAAGTTGCTGTTGCTCAATTA 4613  
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 4614 GCTGATAAGTATACACGTCAGATGTTACATTTTGTGATGACATGATATATCATGATGAT 4673  
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560  
 Db 4674 GAAGGAGATGCATATGTAACGCCTCATATGGGCCATAGTCACTGGATTGGAAGATAGC 4733  
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 4734 CTTTCTGATAAGGAAAAAGTTCCAGCTCAAGCCTATATCAAGAAAGAAAGGTATCTTACT 4793  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600  
 Db 4794 CCATCTCCAGACGAGATGTTAAAGCAATCCACTGGAGATAGTGCAGCACATATTAC 4853  
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 4854 AATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCCGACTTCCATATATGTTGAGCAT 4913  
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsnIle 640  
 Db 4914 ACAGTTGAGTTAAAGCCGTAATTTGATTTATTCCTCAAGATCATTTACCATATATT 4973  
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 4974 AATTTGCTTGGTTTGTATGATGATCACACATCAAGCTCCAAATGGCTATACCTTGGAGAT 5033  
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680

Db 5034 TTGTTTCGCGAGTAAAGTACTAGTAGAACACCTCGACGAACGTCACATTTCTAATGAT 5093  
 Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 5094 GGATGGGCGAATGCCACTGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 5153  
 Qy 701 LysAsnPhelLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 5154 AAGAACTTCAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAAGTCCCT 5213  
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 5214 CAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAGCAGAAAGTTTTCCTTTCGGA 5273  
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsn 760  
 Db 5274 GTAACGATTTCTAGTCTGAAGCCATGCAACAGAACTTAGCTGTTTACGAAATAAT 5333  
 Qy 761 LeuThrLeuGlnIleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 5334 TTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGCAGAAAGCAAAAAATTACTTGGC 5393  
 Qy 781 LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796  
 Db 5394 TTGTTAAAGAGAGTAATCTTCTCATCTGTAAAGTAAAGAAATAAATTAAC 5441  
 RESULT 10  
 LOCUS AE007418/1 10256 bp DNA linear BCT 31-AUG-2001  
 DEFINITION Streptococcus pneumoniae TIGR4 section 101 of 194 of the complete genome.  
 ACCESSION AE007418 AE005672  
 VERSION AE007418.1 GI:14972649  
 KEYWORDS Streptococcus pneumoniae TIGR4  
 SOURCE Streptococcus pneumoniae TIGR4  
 ORGANISM Streptococcus pneumoniae TIGR4  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 10256)  
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.  
 Complete genome sequence of a virulent isolate of Streptococcus pneumoniae  
 Science 293 (5529), 498-506 (2001)  
 JOURNAL Science 293 (5529), 498-506 (2001)  
 MEDLINE 21357209  
 PUBMED 11463916  
 REFERENCE 2 (bases 1 to 10256)  
 Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E., Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.  
 Direct Submission  
 Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 TITLE Location/Qualifiers  
 JOURNAL 1..10256  
 FEATURES  
 source  
 /organism="Streptococcus pneumoniae TIGR4"  
 /mol\_type="genomic DNA"  
 /strain="TIGR4"  
 /db\_xref="taxon:170187"  
 complement(100..702)  
 /gene="SP1170"



```

CDS
complement(100..702)
/gene="SP1170"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK75279.1"
/db_xref="GI:14972650"
/translation="MSVDENEAVNYRFTSDCOLANSIYOSLEKFPDKKNFSGDLIF
TWKSPSIVKSGDYIGRDSOVNLRVIGNIPFNLTNRKSLNNNRGCMGDFPHDF
DIVLDHVAKYAYEKVNNIKYIYELKRALHQLHCEALYFRFNSDDFLEKNYLTWQ
VSKETPFEMDFNFKNISEKIIFERGSKMLNDLKSNYK"
complement(760..1473)
/gene="SP1171"
/complement(760..1473)
/notes="SP1171"
/notes="identified by match to PFAM protein family HMM
PF00702"
/codon_start=1
/transl_table=11
/product="hydrolase, haloacid dehalogenase-like family"
/protein_id="AAK75280.1"
/db_xref="GI:14972651"
/translation="MFYKFLFLDLHTLLDFDAEDVALTQLKEGVADIAQKYY
VPMKALMKLEKKISKQLVNTFRSLFAHFGQKXGSLAQRYQYVLAQOQQTLS
GAHDLDSLTERDYNLYAATNGITAIQTGRLAQSLAPYFNQVFISSLOLQKPDALF
YEKIGQIAGFSKELTIMIGSLTADIQGNNAGIDITWYNPHLENHTQAQTYEVY
SYQDLDCDLNLEKITP"
complement(1724..1843)
/gene="SP1172"
/complement(1724..1843)
/notes="SP1172"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK75281.1"
/db_xref="GI:14972652"
/translation="MLYVVPFYNRSETIVFLNCESIKTDCGAILALETFKN"
complement(1933..2307)
/gene="SP1173"
complement(1933..2307)
/notes="SP1173"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK75282.1"
/db_xref="GI:14972653"
/translation="MESIGLVIVSHSKHIAEGVWELISKVAKDVPIITYVRGTEGGIG
TSFEQVDRVVSNPADTLLAFDLSAKMNLKWTFDSKSIINRVPIVEGAYNAA
LLQAGALSVIOTQLAELEINK"
complement(2729..5188)
/gene="SP1174"
complement(2729..5188)
/notes="SP1174"
/notes="similar to GP:4249624; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75283.1"
/db_xref="GI:14972654"
/translation="WKINKYLAGSVAVLALSVCSYELGRYQAGQDKKSNRVAVID
DQAGQKAEMLTDEVSKEGINAEQIVKITDQGVTSHGHDHYHYNGKVPYDAIIE
ELLMKDPYQLKSDIVNIGKGYIVKNGYVYLKDAHADNITKKEIKRQKER
SHNHSRADNVAARAAQRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKQEL
SASLAAAYVWNGQGRSSSSYNANPAQPLSENHNTVTPYTHQNGENTSSL
LRELAKPLSERHVESDGLIFDPAQITSTRARGVAVPHGNHYHFIPIYQMSLEKRIA
RIIPLRYRSHNVDPSPRPEPSQPTPEPSQAPNPIDEKLVKAVRGKGYGV
FEENGVSYPYIPAKDISAETAGDISHLAKQESLSHKLTKTKDLPSSDREFFNKAYDL
LARIHQDLNKGQVDFEALDNLLERLKVSSDKVLVEDILAFIAPIRPERLGP
NAQITYTDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDYVYTPMTHSHWIKDLSL

```

```

gene
CDS
complement(5346..7754)
/gene="SP1175"
/complement(5346..7754)
/notes="SP1175"
/notes="similar to GP:4249624; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75284.1"
/db_xref="GI:14972655"
/translation="MILSVCSYELGLVQARTVKNNEVSYIDGKQATQKTNLTPTDEV
SKREGINAEQIVKITDQGVTSHGHDHYHYNGKVPYDAIIEELLMKDPYQLKSD
IIVNEKGGYIVKNGYVYLKDAHADNITKKEIKRQKESHQREGGTPRNDGAV
ALARSQRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYIPKNSLSASLAAAEAL
SGRGLNSRTYRQNSDTRTNVPSVSNPGTNTNTNNTNSQAOSQNDISL
LKQLYKPLSRHVESDGLVFDPAQITSTRARGVAVPHGDHYHFIPIYQMSLEKRIA
RIIPLRYRSHNVDPSPRPEPSQPTPEPSQAPNPIDEKLVKAVRGKGYGV
VYFEEKISRYVFAKDLPSFTVKNLESKLSQESVSHLTAKENVAPROEFYDKAG
NLTLEAKHALFENKGRNSFDQALDKLERLNDSTNKEKLVDDLLAFIAPIRPERL
KPSQIYETEDEVRIAQDLADKYTTSDGYIFDEHDIISDEGDYVTPMTHSHWIKGS
LSDEKVAQAQNLIIIPKHDHYNIKFADFDDHTYKAPNGYTTLEDLFATIKYVEHDERP
EHTVEKNGALIIIPKHDHYNIKFADFDDHTYKAPNGYTTLEDLFATIKYVEHDERP
HSNDGKNASHEVLGKKHSDPNKPKFADDERPVEETPAEPPEVQVETEKVEALKEA
EVLLAKVTDSSILKANATETLAGLRNNLTQIIMDNNSIMAEAEKLLALLKGSNFSVSK
EKIN"
complement(8049..9782)
/gene="SP1176"
complement(8049..9782)
/notes="SP1176"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAK75285.1"
/db_xref="GI:14972656"
/translation="WTEMLKGIAGSDGVAVAKAYLLVQPLSPETITVEDTNAEARL
DAALQASQDELVSIREKAVGTIGEEAAQVFDHMLVADPEMISQIKETRAKKVNAE
AGLKEVTDNFIIFEGMEDNPQERAAADIVTKRKLANLILGKLKLPASINEEVIV
IAHDLTPSQADLNKFNKVFNTGGRTSHSAIMARTLEIAIAGLTNNITETVKGOD
ILAVNGITGEVILNPTDEQAEFKAAAGEAYAKQAEWALLKDAQVTAQDKGFELAA
IGTPKDVGVNNGAEAGLYRTFLYMDSQDPTDEQYEAQVLEKMGKPKPVVR
TMDIGDKELPYFDMPHMNPFLGFRALRISISSETGDAMFRTQIRALLASVHGQIRI
MFPVALLKEFEAKAVDEEKANLLAEGVAVADNIQVGMIEIPAAAMLAQDQFAKEV
DFESIGTNDLIQYTHAARMNQSVLYQYINPNSILRLINNVIKAHAEGRKAGMCGE
MAGDQAVLLVGMGLDBFMSQSVLRTSLMKKLDLTAKEEYANRALTECSTMEEV
LELQKEYVNF"
complement(9788..10051)
/gene="SP1177"
complement(9788..10051)
/notes="SP1177"
/notes="similar to GP:4249624; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein Hpr"
/protein_id="AAK75286.1"
/db_xref="GI:14972657"
/translation="MASKDFHVVAETGTIHARPATLLVQTASKFASDITLEYKGSVNL
KSIIMVSLGVCGQADVTIISAGGADDAIAAISSETMEKEGLA"

```

## ORIGIN

## Alignment Scores:

Pred. No.:	Length:	Matches:
4, 61e-229	4163.00	10256
Score:	99.8%	795
Percent Similarity:	Conservative:	0
Best Local Similarity:	Mismatches:	1
Query Match:	Indels:	0
DB:	Gaps:	0

US-09-765-271-56 (1-796) x AB007418 (1-10256)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr 20  
Db 7736 TCTTACGAGTTGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATTCGTGTTTCCTAT 7677  
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
Db 7676 ATAGATGGAAACAAAGCGACGCAAAAACCGAGAAATTTGACTCTCTGATGAGGTAGCAAG 7617  
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
Db 7616 CGTGAAGGAATCAATCGTGAAGCAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 7557  
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
Db 7556 TCACATGGCGACCACTATCATATTATCAATGGTAAGGTTCTCTATGACGCTATCATCAGT 7497  
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
Db 7496 GAAGAATTACTCATGAAGAATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 7437  
QY 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
Db 7436 GTCAAGGTTGATATGTTATCAAGTAGATGGAATACTATGTTTACCTTTAAGCATGCT 7377  
QY 121 AlaHisAlaAspAsnValArgThrLysGluLileAsnArgGlnLysGlnLysHisSer 140  
Db 7376 GCCCAGCGGATAAGCTCGTACAAAAGAGGAATCAATCGACAAAAACAAGAGCATAGT 7317  
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
Db 7316 CAACATCGTGAAGTGGAACTCCAAAGAACGATGTGTGTTGCTTCGCGACGTTCCGAA 7257  
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
Db 7256 GGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATACT 7197  
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlnLeu 200  
Db 7196 GGTGATGCTTATATCGTTCTCATGGAGATCAATACCATTCATTCCTTAAGNATGAGTTA 7137  
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
Db 7136 TCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCCTATCTGTGTCGAGGAATCTGTCAAA 7077  
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
Db 7076 TCAAGAACCTATCGCCGACAAAATAAGCGATAACACTTCAAGAACAACTGGGTACTCTTCT 7017  
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
Db 7016 GTAAGCAATCCAGGAACATAAATACTAACACAGCAACACAGCAACACTAACAGTCNA 6957  
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
Db 6956 GCAAGTCAAAAGTAATGACATTGATAGTCTCTTGAACACAGCTCTACAACTGCGCTTTGAGT 6897  
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
Db 6896 CAACGACATGTGAATCTGATGGCTTGTCTTTGATCCAGCAACAAATCAAGTCAGCAACA 6837  
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
Db 6836 GCTAGAGGTTGTCAGTGCACACGAGATCATACCATCTCATCCCTTACTCTCAATG 6777  
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
Db 6776 TCTGAATTCGAAGACGAATCGCTGATATTATTTCCCTTCGTTATCTGTTCAAAACCATG 6717  
QY 341 ValProAspSerArgProGluGlnProSerProGlnProGluProGluProSerProGly 360  
Db 6716 GTACACAGATTCAAGGCCAGAACCAACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGGC 6657

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
Db 6656 CCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTTGGTTAGTACGCTGGTA 6597  
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
Db 6596 CGAAAAGTTGGGAAGGATATGATTTCGAAGAAAGGGCATCTCTCGTTATGTTCTTTGCG 6537  
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
Db 6536 AAGATTTCACATCTGAACCTGTTAAATACTTGAAGCAAGTTATCAAAACAAGAGAGT 6477  
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
Db 6476 GTTTCACACACTTTAACTGCTGCTAAAAAAGAAATGTTGCTCTCTGCTGACCAAGAAATTTAT 6417  
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
Db 6416 GATAAAGCATATAATCTGTTAACTGAGGCTCATTAAGCCTTGTTGAAAATAAGGTCGT 6357  
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
Db 6356 AATTCTGATTTCCAAAGCTTAGACAAATTAATGAACGCTTGAATGATGAATCAGCTAAT 6297  
QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
Db 6296 AAAGAAAATTTGGTAGATGATTTATGTCATCTCCAGCACCATTACCCATCCAGAGCGA 6237  
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
Db 6236 CTTGCGCAACCAAAATCTCAAAATGAGTATATCTGAAGCGAAGTTCGTTATGCTCAATTA 6177  
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
Db 6176 GCTGATAAGTATACACGTCAGATGGTTTACATTTTGTATGAACATGATATATAATCAGTGAT 6117  
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
Db 6116 GAAGAGATGCATATGTAACGCTCATATGGGCCCATCTGCTGATTTGGAAGATAGC 6057  
QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
Db 6056 CTTTCTGATAGGAAAAAGTTGACCTCAAGCTATATCTAAAGAAAAAGGTATCTCTACCT 5997  
QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600  
Db 5996 CCATCTCCAGCGCAGATGTTAAAGCAAAATCCAACCTGGAGATAGTGCAGCAGCTATTTAC 5937  
QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
Db 5936 AATCGTGTGAAGGGGAAAAACGAATCCACTCGTTCGACTTCCATATATGTTGAGCAT 5877  
QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
Db 5876 ACAGTTGAGTTAAACCGTAAATTTGATTATCTCTCATAGGATCAITTACCATAATATT 5817  
QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
Db 5816 AATTTGCTTGTGTTGATGATCACACATACAAAGCTCCAAAATGGCTATATACCTTGAAGAT 5757  
QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
Db 5756 TTGTTTGGACCAATTAAGTACTACGTAGAACACCTCGACGAACGTCACATCTTAATGAT 5697  
QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
Db 5696 GAATGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTCAAGATCCAAAT 5637  
QY 701 LysAsnPheLysAlaAspGluProValGluThrProAlaGluProGluValPro 720  
Db 5636 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCGAGAGTCCCT 5577  
QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740



```

Db      5576  CAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAGAGCAGAGATTTCCTGCGGAA 5517
QY      741  ValThrAspSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db      5516  GTAACGGATTCCTAGCTGAAAGCCAAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 5457
QY      761  LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db      5456  TTGACTCTTCAAAATATGATAACAATAGTATCATGGCAGAGCAGAGAAATTTACTTGGC 5397
QY      781  LeuLeuLysGlySerAsnProSerValSerLysGluLysIleAsn 796
Db      5396  TTGTTAAAGGAAGTAATCTTCTATCTGTAAGTAAGCAAGAAATAAAC 5349

RESULT 11
AX571763/c
LOCUS   AX571763 349980 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 4982 from Patent WO02077021.
ACCESSION AX571763
VERSION   AX571763.1 GI:26003955
KEYWORDS
SOURCE   Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE
1. Masignani, V., Tettelin, H. and Fraser, C.
   TITLE   Streptococcus pneumoniae proteins and nucleic acids
   JOURNAL Patent: WO 02077021-A 4982 03-OCT-2002;
   Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
   Location/Qualifiers
     1. 349980
        /organism="Streptococcus pneumoniae"
        /mol_type="unassigned DNA"
        /db_xref="taxon:1313"
        /notes="seq 4979 too long: 2,162,598 bases-replaced by
        following seq:-seq 4979: from 0.000.001 to 0.349.980-seq
        4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001
        to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq
        4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001
        to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq
        4986: from 2.100.001 to 2.162.598"

ORIGIN
Alignment Scores:
Pred. No.: 3,84e-227 Length: 349980
Score: 4163.00 Matches: 795
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.95% Indels: 0
DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AX571763 (1-349980)

QY      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
Db      106949 TCTTACGAGTTGGGACTGATCAAGCTAGACGGTTAAGGAAAAATATCGTGTTCCTAT 106890
QY      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      106889 ATAGATGGAACCAAGCGACGCAACCAACCGAGAAATTTGACTCTCTGATGAGGTAGCAAG 106830
QY      41  ArgGluGlyLysAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      106829 CQTGAGGAAATCAATGCTGAGCAATATCGTCATCAAGATTAACAGCAAGCTATGTCAT 106770
QY      61  SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      106769 TCACATGGCGACCACTATCATTTATACATGTTAGGTTTCCTTATGACCTATCATCAT 106710
QY      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

```

```

Db      106709 GAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTGTTAATGAG 106650
QY      101  ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      106649 GTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTTAAGGATGCT 106590
QY      121  AlaHisAlaAspAsnValArgThrLysGluLysLeuIleAsnArgGlnLysGlnGluHisSer 140
Db      106589 GCCCACCGGATTAACGTCGCTACAAAAGAGAGAAATCAATGCACAAAAACAAGACATAGT 106530
QY      141  GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db      106529 CAACATCGTGAAGTGGAACTCCAAAGAAACGATGGTCTGTTGCTTGGCACGTTTCGCA 106470
QY      161  GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      106469 GGACGCTTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 106410
QY      181  GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      106409 GGTGATGCTTATATCGTTCCTCATGGAGATCATTCACATTACATTCCTAAGAATGAGTTA 106350
QY      201  SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      106349 TCAGCTAGCGAGTTGCTGCTGCAGAGCCCTTCTATCTGGTCGAGAAATCTGTCAAT 106290
QY      221  SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      106289 TCAAGAACCTATCGCGCAACAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT 106230
QY      241  ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      106229 GTAAGCAATCCAGGAATCAAAATACTAAACAAAGCAACAAACACAGTAACTAAGTCA 106170
QY      261  AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db      106169 GCAAGTCAAGATTAATGACATTGATGATCTCTTGAACACAGCTCTACAAACTGCTTTGAGT 106110
QY      281  GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      106109 CAACGACATGTAGATCTCTGATGSCCTTGCTTTGATCCAGCACAAATCAACAGTCCAGACA 106050
QY      301  AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      106049 GCTAGAGGTGTTCAGTGCACGAGATCATTAACACTTCATCCCTTACTCTCAATG 105990
QY      321  SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db      105989 TCTGAATTCGAAGACGAATCGCTCGTATATATCCCTTCGTTATCGTTTCAACACCAATTGG 105930
QY      341  ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      105929 GTACCAGATTCAAGGCCAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCCAGGC 105870
QY      361  ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db      105869 CCGCAACCTGCAACCAATCTTAAAAATAGACTCAAAATCTCTTCTTTAGTTCAGTGGTA 105810
QY      381  ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db      105809 CGAAAAGTTGGCAAGGATATGTAATCGAAGAAAAGGGCATCTCTCGTTATGTTCTTGGC 105750
QY      401  LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db      105749 AAAGATTTCACATCTGAAACTGTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 105690
QY      421  ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db      105689 GTTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTCGTGCAGCAAGAAATTTAT 105630
QY      441  AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db      105629 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGAAAAATAAGGGTCTG 105570

```

QY	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480		
Db	105569	AATCTCGATTCCAGCCCTTACAAATTTATTAGAACGCTTGAATGATCGACTAAT	105510		
QY	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500		
Db	105509	AAAGAAAATTGCTAGATGATTTATTGGCATTCTTAGCACCAATTAACCATCCAGACGGA	105450		
QY	501	LeuGlyLysProAsnSerGlnIleGluThrGluAspGluValArgIleAlaGlnLeu	520		
Db	105449	CTTGGCAAAACCAATTTCTCAAATTGATGATCTAGAACGCAAGTTCGTATTGCTCAATTA	105390		
QY	521	AlaAspLysThrThrThrSerAspGlyTyrllePheAspGluHisAspIleIleSerAsp	540		
Db	105389	GCTGATAAGTATACAACGTCAGATGGTTTCAATTTTGTATGACACATATATCATGTGAT	105330		
QY	541	GluGlyAspAlaTyrrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer	560		
Db	105329	GAAGGAGATGCATATGTAAACGCCTCATATGGGCCATAGTCACTGGATTGGAAAAGATAGC	105270		
QY	561	LeuSerAspLysGluLysValAlaAlaGlnAlaTyrrThrLysGluLysGlyIleLeuPro	580		
Db	105269	CTTTCTCATAGGAAAAGTTGCGAGCTCAAGCCCTATCTATAAGAAAAGGTATCTCTACT	105210		
QY	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrr	600		
Db	105209	CCATCTCCAGCAGCATGTTTAAAGCMAATCCAACTGGAGTAGTCAGCAGCTATTAC	105150		
QY	601	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrrMetValGluHis	620		
Db	105149	AATCGTCTGAAAGGGGAAAAACGAATTCCTCACTCGTTCGATCCATATATGGTTGAGCAT	105090		
QY	621	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrrHisAsnIle	640		
Db	105089	ACAGTTGAGGTTTAAAAACGGTAAATTTTGATTTATTCCTCATAGGATCATACCAATAATAT	105030		
QY	641	LysPheAlaTrpPheAspAsnHisThrTyrrLysAlaProAsnGlyTyrrThrLeuGluAsp	660		
Db	105029	AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGCTATACCTTGGAGAT	104970		
QY	661	LeuPheAlaThrIleLysTyrrTyrrValGluHisProAspGluArgProHisSerAsnAsp	680		
Db	104969	TTGTTTGGCAGCATTAAGTACTAGTAGACACCCCTGACGACGTCACATCTAATGAT	104910		
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700		
Db	104909	GGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT	104850		
QY	701	LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro	720		
Db	104849	AAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCCT	104790		
QY	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740		
Db	104789	CAAGTAGAGATGAAAAAGTAGAGGCCCACTCAAGACGACAGAGTTTGTCTTGGAAA	104730		
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760		
Db	104729	GTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAAAATCTAGCTGGTTTACGAAAATAAT	104670		
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780		
Db	104669	TTGACTCTTCAAAATATGATAACAAATAGTATCATGGCAGCAGCAGAAAATTAATTCTCGG	104610		
QY	781	LeuLeuLysGlyLysSerAsnProSerValSerLysGluLysIleAsn	796		
Db	104609	TTGTTTAAAGGAGTAATCTCTCATCTGTAAGTAAGGAAAAATAAAC	104562		
RESULT 12	AE008479/c	10320 bp	DNA	linear	BCT 13-SEP-2001
LOCUS	Streptococcus pneumoniae R6	section 95 of 184	of the complete		
DEFINITION	Streptococcus pneumoniae R6				

RESULT 12  
AE008479/C  
LOCUS  
DEFINITION

```

/gene="spr1058"
/codon_start=1
/transl_table=11
/label=spr1058
/product="Hypothetical protein"
/protein_id="AAK99862.1"
/db_xref="GI:15458680"
/translation="MNLKAVTDFSDKSIINRPVIVEGAYNAALQAGAEISVIQTQ
LAELINK"
complement (2151..2306)
/gene="spr1059"
/codon_start=1
/transl_table=11
/label=spr1059
/product="Hypothetical protein"
/protein_id="AAK99863.1"
/db_xref="GI:15458681"
/translation="MMESIGIGLVSHSHKHAIEGVVELISKVADVPITVVGTEGG
GIGTSPD"
complement (2542..2613)
/rpt_family="RUP element"
/rpt_type="dispersed"
/rpt_unit="rupA"
complement (2725..5292)
/gene="phpA"
/notes="synonym: spr1060"
complement (2725..5292)
/gene="phpA"
/codon_start=1
/transl_table=11
/label=spr1060
/product="Histidine Motif-Containing protein"
/protein_id="AAK99864.1"
/db_xref="GI:15458682"
/translation="MNOIYLRKEERMKINKYLAGSVATLVLSVAYELGLHQATVK
ENRVSVIDGKATQKTENITPDEVSKREGINAEQIVIKITDOGYVTSBGDHYHYNG
KVPYDAIIEELMKDPNQKDEDIIEIKGGVIVKDGKYVYLKDAHADNVRTK
EINROKQHSORHGGTFRNDGAVALARSGQRTYDDGYIFNADIIEDTDAIYV
HGDIHYIPKNEISASLAAKAFSLSGNLSNRTYRQNSDNTSRNTWVPSVNP
TTNTNNSNTSQSOSNDISLLQYKPLSRHVESDGLIFDPAQITSRNG
VAVPHGDIHYFIPYQSPLSEKLARIIPLYRSNHWVPSRPEQSPQSTPPSPSP
QAPNPQPPSPNIDELKVEAVKVGQGVFEENGVPYIPAKDISAETAGIDSKL
AKQESHLKAGKTLPLSSDREFYNKAYDILLARIHODLONKROVDFEALNLER
LKDVSSDKVLVDIILAFAPIRPERLGPKNQAITYTDEIQVAKLAGYITTEDGYI
NTEAKGAEALYNRVAAKVPDRMPINQITVEVKNGLIIPHYDHYNIKEWFEDE
GLYEAQKGSLELLARVYVVEHPNPHSDNGFNASDHVQRNKGQADNTQTEKP
NEEKPTKEPEBETPREKPEKSPSPKTEPEEPEEPEEPEEPEEPEEPEEPEEPEE
KLREABDLGKQNPFIKSNAKETLTGLKNLLFGTDNNTIMAEAEKLLALLKESK"
complement (5417..7903)
/gene="phtA"
/notes="synonym: spr1061"
complement (5417..7903)
/gene="phtA"
/codon_start=1
/transl_table=11
/label=spr1061
/product="Pneumococcal histidine triad protein A
precursor"
/protein_id="AAK99865.1"
/db_xref="GI:15458683"
/translation="MQLEISNRKYSVMKINKYLVGSRAALLVSCVSELGLYQARTV
KENNRVSYIDGKATQKTENITPDEVSKREGINAEQIVIKITDOGYVTSBGDHYHYN
KVPYDAIIEELMKDPNQKDEDIIEIKGGVIVKDGKYVYLKDAHADNVRTK
KEEINROKQHSORHGGTFRNDGAVALARSGQRTYDDGYIFNADIIEDTDAIYV
PHGDIHYIPKNEISASLAAKAFSLSGNLSNRTYRQNSDNTSRNTWVPSVNP
GTNTNNSNTSQSOSNDISLLQYKPLSRHVESDGLIFDPAQITSRNG
GVAVPHGDIHYFIPYQSPLSEKLARIIPLYRSNHWVPSRPEQSPQSTPPSPSP
POPAPNLKIDSNSLVSQLVRKVGQGVFEENGISRYVFAKDLPSFTVKNLESLSKQ
ESVSHTLKAKENVAPROQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLERLND
ESTNKEKLVDDLLAFAPITPHERLGPKNQIETVEDEVRIQAADKYVTSBGYIFDE

```

```

HDIISDEGDYVTPMHSHWIGKDSLCKEKVAQAQYTKKGILPPSPDADKANPT
GDSAAAIYNRVYKGERIIPVRLPYVVEHTVEVKNGLIIPKHQVHNIKFAWDDHTY
KAPNGYTLIEDLFIATIKYVVEHDPDRPHSDNGWGNASEHVLGKDHSEDPKNKPKADEE
PVETPARPEVPEQVETEKVQAQLKEAEVILAKVTDSLSKANATETILAGLNNLITQIOM
DNNSIMAEAEKLLALLKSNFSSVSKEKIN"
complement (8122..9855)
/gene="ptsI"
/notes="synonym: spr1062"
complement (8122..9855)
/gene="ptsI"
/EC_number="2.7.3.9"
/codon_start=1
/transl_table=11
/label=spr1062
/product="PTS enzyme I"
/protein_id="AAK99866.1"
/db_xref="GI:15458684"
/translation="WTEMLKGIAASDGVAVAKAYLLVQPDLSFETITVEDTNAEERL
DAIQAODELSVIREKAVGTGEEAAQVFDALHMLVADPEMISOIKETIRAKVNAE
AGLEVTDMFTITFEQMEDNPMQBERAADIRDTKRVLANLLGKLPNPASINERVIJ
IAHDLTPSDTAQLDKNFVKAFVTNIGRTSHSAIMARTLEIAAVLTNNITETIYKDG
ILAVNGITGEVINPTDQAAEFKAAGEAYAKQAEKALLKDAQTVTADKHFEAAN
IGTPKQVEGVNNGAABVGLYRTEFLMDSQDPTEDQYEAQVLEGMNGKPVVVR
TMDIGCKELPYFDMPEHMPFLGFRALRISISSETGDAMFRTQIRALIRASVHGQIRI
MFMWALLKEFRAAKAVDEEKANLLAEVAVADNIQVGINETIPAAAMLADQFAKEV
DFFSIGTNDLIQYTMADRMNEQVSYLYQVYNPSILRLINNVIKAAHAEGKAWGCGE
MAGDQAVPLLVGMGLDFSMATSVLTRLSMKKLDTKAKMEEVANRALTECSTMEEV
LELQKEYVNFQ"
complement (9861..10124)
/gene="ptsH"
/notes="synonym: spr1063"
complement (9861..10124)
/gene="ptsH"
/codon_start=1
/transl_table=11
/label=spr1063
/product="Histidine-containing phosphocarrier protein of
the PTS"
/protein_id="AAK99867.1"
/db_xref="GI:15458685"
/translation="MASKDFHVVAETGIIHARPATLLVQVTSKFPASDITLEYKGSVNL
KSMGVMSLVGGQADVITSAEGADADDAIAAISETMEKEGLA"
ORIGIN
Alignment Scores:
Pred. No.: 7,87e-229 Length: 10320
Score: 4159.00 Matches: 794
Percent Similarity: 99.75% Conservative: 0
Best Local Similarity: 99.75% Mismatches: 2
Query Match: 99.86% Indels: 0
DB: 1 Gaps: 0
US-09-765-271-56 (1-796) x AE008479 (1-10320)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
Db 7807 TCTTACGAGTTGGGACTGTATCAAGCTAGACGGTTAAGGAAAAATAATCGTGTTCCTAT 7748
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 7747 ATAGATGGGAAACAAACGCGACGCAAAAAACGGAGAAATTTGACTCTCTAGGTTAGCAAG 7688
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGluGlyTyrValThr 60
Db 7687 CGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGGCTATGTCTACT 7628
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 7627 TCACATGCGCACCATCATCATTTTACAAATGGTAAAGTTTCCTTATGATCGCGATTTTCAGT 7568
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 7567 GAAGAACTACTCATGAAAGATCCAAACTATAAGCTAAAAAGATGAGGATATTGTTTAATGAG 7508

```

QY	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAenAspGluSerThrAsn	480
Db	6427	AAITCTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGTTGAATGATGAATCGACTAAT	6368
QY	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
Db	6367	AAAGAAAAATTGGTAGATGATTTATTGGCATTCCTTAGCACCATTACCCATCCAGACGGA	6308
QY	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520
Db	6307	CTTGGCAAGCCAAATCTCTCAAAATTCAGTATCTAGTGAAGACGAAGTTCGTATTGCTCAATTA	6248
QY	521	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp	540
Db	6247	GCTGATPAAGTATACAACGTCAGATGGTTTACATTTTGTATGAACATGATATATATCAGTGAAT	6188
QY	541	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrioIleGlyLysAspSer	560
Db	6187	GAAGGAGATGCATATGTAACGGCTCATATGGGCCCATAGTCACTGGATTGGAAACACAGC	6128
QY	561	LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro	580
Db	6127	CTTTCTGTAAAGGAAAGAGTTGCAGCTCAAGCCTACTATAAGAAAAAGGTATCTACCT	6068
QY	581	ProSerProAspAlaAspValLysAlaAenProThrGlyAspSerAlaAlaIleTyr	600
Db	6067	CCATCTCCAGCCGACAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTTTAC	6008
QY	601	AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
Db	6007	AATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTTCGACTCTCCATATATGTTGAGCAT	5948
QY	621	ThrValGlnValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
Db	5947	ACAGTTGAGGTATAAAACGGTAATTTGATTAATTCCTCATAGGATCATACCATATAATTT	5888
QY	641	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
Db	5887	AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGGAAGAT	5828
QY	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	5827	TTGTTTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAACGTCCACATCTAATGAT	5768
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
Db	5767	GGATGGGCAATGCCAGTGACATCTGTTAGGCAAGAAAGACCCACAGTGAAGATCCAAAT	5708
QY	701	LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro	720
Db	5707	ANGAATCTCAAAGCGATGAAGACCCAGTACAGAGGAAACACCTGCTGAGCCAGAAAGTCCCT	5648
QY	721	GlnValGluThrGluLysValGluAlaGlnIleLysGluAlaGluValLeuLeuAlaLys	740
Db	5647	CAAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAAAGTTTGTCTTGGAAA	5588
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
Db	5587	GTAACGGATCTAGTCTGAAAGCCCAUGCAACAGAACTCTAGCTGGTTTACGAAATAAT	5528
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
Db	5527	TTGACTCTTCAAATTTATGATAACAAATAGTATCATGCGCAGAAAGCAGAAAAATTACTTGC	5468
QY	781	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
Db	5467	TTGTTAAAAGGAAGTAATCTCTCATCTGTAAAGTAAGGAAAAAATAAAC	5420
RESULT 13			
SPNEU1915			
LOCUS			
DEFINITION			
ACCESSION			
AL449937			
20035 bp DNA linear HTG 11-JUL-2001			
Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.			

RESULT 13	
SPNEUL915	
LOCUS	
DEFINITION	
ACCESSION	

QY	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrValTyrValTyrLeuLysAspAla	120
DB	7507	GTCAGGGTGGATATGTTATCAAGTAGATGGAAATATATGTTTACCTTAAGGATGCT	7448
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluLeuLeuAsnArgGlnLysGlnUuHisSer	140
DB	7447	GCCACCGGATAACGTCGTACAAAGAGGAAATCAATCGACAAAAACAAGAACATAGT	7388
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160
DB	7387	CAACATCGTGAAGGTGGGACTCCAGAAACGATGTGTGTTCCTGGACGTCGTCGAA	7328
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
DB	7327	GGACGCTATCTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGTACT	7268
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProIleProIleAsnGluLeu	200
DB	7267	GGTGATGCTTATATGCTCTCATGGAGATCAATTACCATTCAATTCCTTAAGAAATGAGTTA	7208
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
DB	7207	TCAGCTAGCGAGTGGCTGCTGCAGAAAGCTTCCTATCTGGTCGGGAAATCTGTCAAT	7148
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
DB	7147	TCAAGAACCCTATGCCGACAAATAGCGATAACACATTCAGAAACAACTGGGTACTTCT	7088
QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
DB	7087	GTAAGCAATCCAGGAATCAAAATACTTAACACAGCAACAACAGCAACACTAACAGTCAA	7028
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
DB	7027	GCAAGTCAAAAGTAAATGATGTAGTGTCTTGAAACAGCTCTACAACTGCTCTTGAGT	6968
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnLeuThrSerArgThr	300
DB	6967	CAACGACATGTAGAACTGTAGTGGCTGTGTTTCGACCCAGCGCAATCAAAAGTCGAACA	6908
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
DB	6907	GCTAGAGGTGTGCAGTGCCACACGGAGATCAATTACCCTTCATCCCTTACTCTCAATG	6848
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
DB	6847	TCTGAATTGGGAAGACGAATCGCTCGTATATATCCCTTCGTTATCGTTCMAACCAATGG	6788
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
DB	6787	GTAACAGATTCAAGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC	6728
QY	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerIleuValSerGlnLeuVal	380
DB	6727	CCGCAACCTGCACCAAAATCTTAAAAATAGACTCAAAATCTCTTTTGGTTAGTACAGCTGGTA	6668
QY	381	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
DB	6667	CGAAAAGTGGGGAAGGATATGATTCGAGAAAGGCACTCTCTCGTTATGCTTTTGGC	6608
QY	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
DB	6607	AAAGATTTACCATCTGAACCTGTAAAAATCTTGAAGTAAAGTTTATCAAAACAAGAGAGT	6548
QY	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
DB	6547	GTTTCCACACATTTAACTGCTAAAAAAGAAAATGTTGCTCTCTCGTGACCAAGAAATTTAT	6488
QY	441	AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
DB	6487	GATAAAGCATATTAATCTGTTAACTAGAGCTCATAAAGCCTTGTTGAAAAATAAGGGTCGT	6428

```

VERSION AL49937.1 GI:11545162
KEYWORDS HTG, HTGS PHASE2.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.
and Garcia-Bustos, J.F.
TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
MEDLINE 21335329
PUBMED 11442348
REFERENCE 2
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and
Garcia-Bustos, J.F.
TITLE Direct Submission.
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
location/Qualifiers
1..20035
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
ORIGIN
Alignment Scores:
Pred. No.: 8,97e-226 Length: 20035
Score: 4112.00 Matches: 787
Percent Similarity: 98.75% Conservative: 0
Best Local Similarity: 98.75% Mismatches: 9
Query Match: 98.73% Indels: 1
DB: 2 Gaps: 0
US-09-765-271-56 (1-796) x SPNEU1915 (1-20035)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 17071 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTTCCTAT 17130
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 17131 ATAGATGGAAACCAACGGACGCAAAACGGAGAAATTTGACTCTCGATGAGGTAGCAAG 17190
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 17191 CQTGAAGGAATCAATCTCGACCAATCGTCATCAAGTAAACAGCCAGCGCTATGTCAT 17250
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 17251 TCACATGGCGGACCATCATATTACATGTTAAGTGGTAAAGTTCCTATGATCGATTTTCAGT 17310
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValIleGlu 100
Db 17311 GAAGAAATTTACTCATGAAAGATCAAACTATAAGTAAAGATGAGGATATTGTTAATGAG 17370
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 17371 GTCAGGGTGGATATCTTATCAAGGTAGATGAGAAATATACTATGTTTACCTTAAGGATGT 17430
QY 121 AlaHisAlaAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140

```





Db 301 GAAGAAATTCCTGATGAAGATCCAACTACAGCTTAAAGGACGAGGATATTATCATGATGAA 360  
 Qy VallysGlyGlyTyrValValleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 361 ATCAAGGGTGGTTATGTGATTAAGTAGATGAAGAAATATCTATGTTTACCTTAAGGATGCA 420  
 Qy 121 AlaHisAlaAspAsnValArgThrLysGluLysGluLysAsnArgGlnLysGlnGluHisSer 140  
 Db 421 GCTCATGCGGATAATGTCGTACAAAAGAGAAATCAATCAACAAAACAAAGAGCATAGT 480  
 Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 481 CAACATCGTGAAGTGGGACTCCAAAGAACCATGGCTGTGCTTGGCAGCGTTCCAG 540  
 Qy 161 GlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuGluAspThr 180  
 Db 541 GGACGCTACACCAAGATGATGGTTATATCTTAAATGCTTCGTATCATTTGAAGATACT 600  
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 601 GGTGATGCTTATATCGTTCCTCACCGCGACCATTTACCATTTACATTCCTAAGAATGAGTTA 660  
 Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 661 TCAGCTAGCGAGTAGCTGCTGCAAAAGCCTTCTATCTGCTGGGGAATCTGTCAAA 720  
 Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 Db 721 TCAAGAACCTATCGCGACAAATATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT 780  
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 781 GTAAGCAATCCAGCAACTACAAATACTAACACAAAGCAACACAGCAACACTAACAGTCAA 840  
 Qy 261 AlaSerGlnSerAsnAspLeuSerLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 841 GCAAGTCAAAGTAAATGACATTTGATAGTCTCTTGAAACAGCTCTACAACTGGCTTTGAGT 900  
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 901 CAACGCGCATGTAGAAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCAAGTCGAACC 960  
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProLeuArgTyrArgSerAsnHisTyr 340  
 Db 961 GCCAATGGTGTGCTGTACCGCACGAGACCATTTATCACTTTATCTCTTATTCACAACTG 1020  
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
 Db 1021 TCACCTTTGGAAAGAAATTTGGCTCGTATTTATTCCTCTCGTTATCGTTCAAACCATGG 1080  
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1081 GTACAGATTCAAGACCAGAACCAACAGTCCACATCCAGCTCCGGAACCTAGTCCAACT 1140  
 Qy 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374  
 Db 1141 CGCGAACCTGCACCAAAATCTCAACAGCTCCAGCAATCCAAATTCAT-----GAGAAA 1194  
 Qy 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394  
 Db 1195 TGGTCAAAGAACGCTGTTGCAAAAGTAGCGGATGTTATGTTCTTTGAGGAGAATGGAGTT 1254  
 Qy 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414  
 Db 1255 CCTCGTTATATCCAGCCCAAGGATCTTTACAGCAAGAAACAGCAGCAGGATTCATGATAGCAA 1314  
 Qy 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434  
 Db 1315 CTGGCCAAAGCAAGAGTTTATCTCATAGCTAGAGCTAGAAACCTGACCTCCCATCT 1374  
 Qy 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454

Db 1375 AGTGATAGAGAATTTTCAATAAAGCTTATGACTTACTAGCAAGAATTCATCAAGATTTA 1434  
 Qy 455 Phe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474  
 Db 1435 CTTGATAATAAAGTCGACAAAGTTGATTTTTCAGGCTTTGGATAACCTGTTGGAAAGCATC 1494  
 Qy 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494  
 Db 1495 AAGGATGTCCTCAAGTATAAAGTCAAGTTAGTGTGATGATATTCCTTGCCTCTTACGTCG 1554  
 Qy 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514  
 Db 1555 ATTCTCATCCAGAACGTTTAGGAAACCAATCGCAAAATTACCTACACTGATGATGAG 1614  
 Qy 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsnGlu 534  
 Db 1615 ATTCAAGTAGCCAAAGTTGGCAGGCAAGTACACACACAGACGCTTATATCTTTGATCCT 1674  
 Qy 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554  
 Db 1675 CGTGATATAACCAAGTATGAGGGGATGCTCTATGTAACCTCCACATATGACCCATAGCCAC 1734  
 Qy 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574  
 Db 1735 TGGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCTAAA 1794  
 Qy 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594  
 Db 1795 GAGAAAGTTTGACCCCTCTCTCGACAGACCATCAGGATTCAGAAATACTGAGGCCAAA 1854  
 Qy 595 SerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeu 614  
 Db 1855 GGAGCAGAAAGCTATCTACAAACCGCTGAAAGCAGCTAAGAAAGGTGCCACTTGATCGTATG 1914  
 Qy 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlelleProHisLys 634  
 Db 1915 CCTTACAATCTTCAATATACTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1974  
 Qy 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn 654  
 Db 1975 GACCATTTACCATTAACATCAAAATTTGAGTGTGTGAGCAAGGCTTTATGAGGCACCTAAG 2034  
 Qy 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674  
 Db 2035 GGGTATAGTCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAA 2094  
 Qy 675 ArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694  
 Db 2095 CGTCCGATTCAGATAATGGTTTGTGAACGCTAGCGACCATGTTCAAAGAAACAAAAT 2154  
 Qy 695 HisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708  
 Db 2155 GGTCAAGCTGATFACCAATCAAAACGGAACCAACCAAGAGGAGAAACCTCAGACAGAAA 2214  
 Qy 709 ProValGluGluThr----- 713  
 Db 2215 CCTGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAAACGAGAAACCTCAGAGTCTCCAAA 2274  
 Qy 714 -----ProAlaGluProGluValPro 720  
 Db 2275 CCAACAGAGGAACCAAGAAAGATTCACAGAAAGATCACCAGAGGAATCAGAGAACT 2334  
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2335 CAGGTCGAGACTGAAAGGTTAAAGAAACCTGAGAGAGGCTGAAGATTTACTTGGAAAA 2394  
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2395 ATCCAGAAATCAATATCAAGTCCAATGCCAAAGAGACTCTCACAGGATTAATAAT 2454  
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780  
 Db 2455 TTACTATTGGCAGCCAGGCAACAACTACTATTATGCGAGAACTGAAAACTATTGGCT 2514

QY	781 LeuLeuLysGlySer 785	2515 TTATTAAGGAGAGT 2529	5253 bp	DNA	linear	PAT 17-JUL-2003
QY	BD263585	BD263585	Novel streptococcus antigens.			
Db	LOCUS	BD263585	Novel streptococcus antigens.			
	DEFINITION	BD263585	Novel streptococcus antigens.			
	ACCESSION	BD263585.1	GI:33073353			
	VERSION	JP 2002533123-A/2.				
	KEYWORDS	unidentified				
	SOURCE	unclassified.				
	ORGANISM	1 (bases 1 to 2523)				
	REFERENCE	Hamel,J., Brodeur,B.R., Pineau,I., Martin,D., Rioux,C. and				
	AUTHORS	Charland,N.				
	TITLE	Novel streptococcus antigens				
	JOURNAL	Patent: JP 2002533123-A 2 08-OCT-2002;				
	COMMENT	SHIRE BIOCHEM INC				
		OS S. pneumoniae				
		PN JP 2002533123-A/2				
		PD 08-OCT-2002				
		PF 20-DEC-1999	JP 2000591190			
		PR 23-DEC-1998	US 60/113800			
		PI JOSEF HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI				
		CLEMENT RIOUX.				
		PI NATHALIE CHARLAND				
		PC C12N15/09, A61K39/00, A61K39/39, A61P11/00, A61P25/00,				
		PC A61P27/16,				
		PC A61P31/04, C07K14/315, C07K19/00, C12N1/15, C12N1/19, C12N1/21, PC				
		C12N5/10,				
		PC C12P21/02, C12N15/00, C12N5/00				
		CC Coding region of BVH-11 gene				
		FH Key	Location/Qualifiers			
		FT CDS	(1)..(2520).			
	FEATURES	Location/Qualifiers				
	source	1..2523				
		/organism="unidentified"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:32644"				
	ORIGIN					
	Alignment Scores:					
	Pred. No.:	1,250-175	Length:	2523		
	Score:	3218.00	Matches:	616		
	Percent Similarity:	82.95%	Conservative:	65		
	Best Local Similarity:	75.03%	Mismatches:	102		
	Query Match:	77.26%	Indels:	38		
	DB:	6	Gaps:	3		
	US-09-765-271-56 (1-796) x BD263585 (1-2523)					
QY	1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20					
Db	61 GCTTATGACTAGTTTGTCATCAAGCTCAACTGTAAGAAAGAAATAAATCGTGTTCCTAT 120					
QY	21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40					
Db	121 ATAGATGGAAACCAAGCGACGCAAAACCGGAGAAATTTGACCTCGTATGAGTTAGCAAG 180					
QY	41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60					
Db	181 CGTGAAGGAATCAACCGCGACAAATCGTCACTCAAGATTACGGATCAAGTTATGTGACC 240					
QY	61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80					
Db	241 TCTCATGAGACCATATCATCTACTATATATGCAAGGTCCTTATGATGCCATCATCACT 300					
QY	81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100					
Db	301 GAAGAGCTCCTCATGAAGATCCGAATATTCAGTTTGAAGATTTCAGCAATTTGTCATGAA 360					



QY 455 Phe\*\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474  
 Db 1435 CTTGTATAATAAAGGTCGACAGTTGATTTTGGGCTTTGGATAACCTGTTGGACGACTC 1494  
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAlaPro 494  
 Db 1495 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTCTTGCCCTCTTAGCTCG 1554  
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluThrThrGluAspGlu 514  
 Db 1555 ATTGCTCATCCAGAACGTTTAGAAGAACCAATGCGCAATTAACCTACACTGATGAG 1614  
 QY 515 ValArgIleAlaGlnLeuAlaAspLysThrThrSerAspGlyTyrIlePheAspGlu 534  
 Db 1615 ATTCAGTAGCCAACTTGGCAGCAAGTACACACAGAGACGGTATATCTTTGATCCT 1674  
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554  
 Db 1675 CGTGATATACCAAGTATGAGGGGATGCTGTATGTAACCTCACATATGACCATAGCCAC 1734  
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574  
 Db 1735 TGGATTAATAAAGATAGTTTCTGAAGCTGAGAGCGGCGAGCCGCTTATGCTAAA 1794  
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594  
 Db 1795 GAGAAAGGTTTGAACCTCTTCGACAGACCATCAGGATTTCAGGAATACTGAGGCAAAA 1854  
 QY 595 SerAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614  
 Db 1855 GGACAGAGAGTATCTCAACCGCGTGAAGACAGCTAAGAGGTGCCACTTGTATGCTATG 1914  
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634  
 Db 1915 CCTTACAATCTTCAATATCTGTAGAGTCAAAACGGTAGTTTAACTACATCCTCATAT 1974  
 QY 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspHisThrTyrLysAlaProAsn 654  
 Db 1975 GACCATTTACCATATACATCAATTTTGAAGTTTGAAGGCTTTTATGAGGACCTTAAG 2034  
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674  
 Db 2035 GGGTATCTCTTGGAGTCTTTGGCCACTGTCAAGTACTATGTGCAACATCCAAACGAA 2094  
 QY 675 ArgProHisSerAsnAspGlyTyrPylAsnAlaSerGlyAlaValLeuGlyLysLysAsp 694  
 Db 2095 CGTCCGATTTCAGATAATGTTTGTGTAACGCTAGGACCATGTTCAAGAAACAAAAAT 2154  
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp 706  
 Db 2155 GGTCAAGCTGATACCAATCAAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAA 2214  
 QY 706 ----- 706  
 Db 2215 CCTGAGGAAGAACCCCTCGAAGAGAAACCAAAAGCGAGAAACAGAGTCTCCAAAA 2274  
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724  
 Db 2275 CCAACAGAGAGAACCAAGAGAAATCACCAGAGGATCAGAGAACCTCAGGTCGAGACT 2334  
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 Db 2335 GAAAGGTTTGAAGAAACCAAGTGAAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2394  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 Db 2395 ATTATCAAGTCCATGCCAAAGAGACTCTCAGCAGGATTAATAAATTAATTTATTTGCG 2454  
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 Db 2455 ACCCAGGACAAACAATACTATTATGCGAGAGCTGAAAAACTATTGGCTTTTATTAAAGGAG 2514  
 QY 785 Ser 785

Db 2515 AGT 2517  
 RESULT 16  
 AX343072  
 LOCUS AX343072  
 DEFINITION Sequence 3 from Patent WO0198334.  
 ACCESSION AX343072  
 VERSION AX343072.1 GI:18152270  
 KEYWORDS  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 REFERENCE 1  
 AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.  
 TITLE Streptococcus antigens  
 JOURNAL Patent: WO 0198334-A 3 27-DEC-2001;  
 SHIRE BIOCHEM INC. (CA)  
 FEATURES  
 Location/Qualifiers  
 source  
 1..2523  
 /organism="unidentified"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32644"  
 /note="BVH-11"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,25e-175 Length: 2523  
 Score: 3218.00 Matches: 616  
 Percent Similarity: 82.95% Conservative: 65  
 Best Local Similarity: 75.03% Mismatches: 102  
 Query Match: 77.26% Indels: 38  
 DB: 6 Gaps: 3  
 US-09-765-271-56 (1-796) x AX343072 (1-2523)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 61 GCTTATGAACGTAGTTTGCATCAAGCTCAAACTGTAAGAAAAATAATCGTGTTCCTAT 120  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 121 ATAGATGGAAAAAAGCGAGCGAAAAACGGAATTTGACTCTGATGAGTTAGTCAAG 180  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 181 CGTGAAGGAATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGTTTATGTGACC 240  
 QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 241 TCTCATGGAGACCATTTATCTACTATAATGGAAGTCCCTTATGATGCCATCATCAGT 300  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 301 GAAGAGCTCTCATGAAGATCCGAATATCATGTTAAGGATTCAGACATTGTCAATGAA 360  
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 361 ATCAAGGGTGGTTATGTCATTAAAGTAAACGGTAATACTAATGTTTACCTTAAAGGATGA 420  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db 421 GCTCATGCGGATAATGTCCTGACAAAAGAGAAATCAATCGCAAAAAACAAGACATAGT 480  
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 481 CAGCATCTGAGGAGGAGGACTTCAGCAACGATGTTGGGTAGCCCTTTCACGTTTACAG 540  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 541 GGACGCTACACACAGATGATGTTTATCTTCAATGATCTGATCATCATGAGATACG 600  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

Novel streptococcus antigens  
Patent: JP 200253123-A 7 08-OCT-2002;

SHIRE BIOCHEM INC  
 OS S. pneumoniae  
 PN JP 2002533123-A/7  
 PD 08-OCT-2002  
 PF 20-DEC-1999 JP 2000591190  
 PR 23-DEC-1998 US 60/113800  
 PI JOSEF HAMEL, BERNARD R BRODEUR, ISABELLE PINEAU, DENIS MARTIN, PI CLEMENT RIOUX,  
 PI NATHALIE CHARLAND  
 PC C12N15/09, A61K39/00, A61K39/09, A61K39/39, A61P11/00, A61P25/00,  
 PC A61P27/16,  
 PC A61P31/04, C07K14/315, C07K19/00, C12N1/15, C12N1/19, C12N1/21, PC C12N5/10,  
 PC C12P21/02, C12N15/00, C12N5/00  
 CC Novel streptococcus antigens  
 FH Key Location/Qualifiers  
 FT source 1..2647  
 FT Location/Qualifiers  
 FT /organism='S. pneumoniae'.  
 FT 1..2647  
 FT /organism='unidentified'  
 FT /mol\_type='genomic DNA'  
 FT /db\_xref='taxon:32644'

ORIGIN

Alignment Scores:  
 Pred. NO.: 1.33e-175 Length: 2647  
 Score: 3218.00 Matches: 616  
 Percent Similarity: 82.95% Conservative: 65  
 Best Local Similarity: 75.03% Mismatches: 102  
 Query Match: 77.26% Indels: 38  
 DB: 6 Gaps: 3

US-09-765-271-56 (1-796) x BD263590 (1-2647)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db :: 164  
 105 GCTTATGAACGTAGTTTGCATCAAGCTCAAGCTCAAACTGTAAGAAAGAAATAATCGTGTTCCTAT 164

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db :: 165  
 165 ATAGATGGAAACACAGCGACGCAAAACCGAGAAATTTGACTCTCTGATGAGTTAGCAAG 224

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db :: 225  
 225 CGTGAAGGAATCAACGCCGCAACAAATCGTCATCAAGATTACGATCAAGTTATGTGACC 284

QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80  
 Db :: 285  
 285 TCTCATGGAGACCATTTATCAATTACTATATTAATGGCAAGTCCCTTATGATGCCATCATCAGT 344

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluValAsnGlu 100  
 Db :: 345  
 345 GAAGACCTCTCATGAAGAGATCCGAATTAATCAAGTTGAGGATTCAGACATTTGCAATGAA 404

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120  
 Db :: 405  
 405 ATCAAGGGTGGTTATGTCATTAAAGGTAACGGTAATACTATGTTTACCTTAAGGATGCA 464

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db :: 465  
 465 GTCATGCGGATAATGTCGTAACAAAGAGAAATAATCAATCGGCAAAACAAACAGACATAGT 524

QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db :: 525  
 525 CAGCATCGTAAGAGGAGACTTCAGCAAAACGATGGTGGCGTAGCCCTTTCAGCTTCACAG 584

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db :: 585  
 585 GCAGCGCTACACCACAGATGATGGTTATATCTTCAATGCACTGATATCATCGAAGATACG 644

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db :: 179

645 GCGGATGCCCTATATCGTTCCTCATGGAGATCATTTACCATTCCTTAAGAATGAGTTA 704

QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db :: 705  
 705 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCCTATCTGGTCGGGAAATCTGCAAAAT 764

QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 Db :: 765  
 765 TTAAGAACCTATCGCCGACAAAATAGCGATAACACTCCAAGAAACAACTGGGTGCTCTCT 824

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db :: 825  
 825 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACAAACAGCAACACTAACAGTCAA 884

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db :: 885  
 885 GCAAGTCAAGATAATGACATTGATAGTCTCTTGAACAGCTCTACAACTGCTTTGAGT 944

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db :: 945  
 945 CAACGCCATGTAGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCAAGTCAAGC 1004

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db :: 1005  
 1005 GCCAGAGGTGAGCTGCTCCCTCATGGTAACCATTAACCACTTTATCCCTTATGAACAAATG 1064

QY 321 SerGluLeuGluGluArgIleAlaAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
 Db :: 1065  
 1065 TCTGAATTGGAAACCAAGATTTGCTCGTATATTATTCCTTCCTTCGTTATCGTTCAACATTTG 1124

QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db :: 1125  
 1125 GTACCAAGATTCAGACCAAGCAAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCAAGT 1184

QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374  
 Db :: 1185  
 1185 CCGCAACTCGACCAAAATCTCTCAACAGCTCCCAAGCAATCCAATTGAT-----GAGAAA 1238

QY 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394  
 Db :: 1239  
 1239 TTGGTCAAGAACGCTGTTTCAAAAGTAGGCGATGTTTATGTTCTTTGAGGAGAAATGGAGTT 1298

QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414  
 Db :: 1299  
 1299 TCTCGTTATATCCAGCAAGAAATCTTTCAGCAAGAACACAGCAGCAGCATTTGATGCAAA 1358

QY 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434  
 Db :: 1359  
 1359 CTGGCCACAGCAAGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTGACCTCCCATCT 1418

QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454  
 Db :: 1419  
 1419 AGTGATCGAGAAATTTTACAATAAGCTTATGACTTACTAGCAGAAATTCACCAAGATTTA 1478

QY 455 Phe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArgLeu 474  
 Db :: 1479  
 1479 CTTGATAATAAAGGTTCGACAGTGTGATTTTGAAGCTTTTCGATAAACCTGTTGGAACGACTC 1538

QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAlaPro 494  
 Db :: 1539  
 1539 AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGGATGATATTTCTTCCCTTCTTAGCTCG 1598

QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514  
 Db :: 1599  
 1599 ATTGCTCATCCAGACGTTTAGAAAAACCAATTCGCAANTTACCTACTACTGATGAG 1658

QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534  
 Db :: 1659  
 1659 ATTCAAGTAGCAAGTTGGCAGGCAAGTACACACAGAAAGCGGTTATATCTTTGATCTCT 1718

QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554  
 Db :: 1719  
 1719 COTGATAAACCAAGTGTAGGGGGATGCCTATGTAACTCCACATATGACCATGACCCATGAC 1778



QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 945 CAACGCCATGTAGAATCTGATGCCCTTATTTTCGACCCAGCGCAATCACAGTGAACC 1004  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 1005 GCCAGAGTGTAGCTGCTCCCTCATGTAACCAATTAACCACTTTATCCCTTATGAACAAATG 1064  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuAlaGlyTyrArgSerAsnHisTrp 340  
 Db 1065 TCTGAATTGGAAACAAAGAAATGCTCGTATATTATCCCTTCGTTATCGTTCAACCAATTGG 1124  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1125 GTACCAAGTTCACAGACCAAGAACCAAGTCCACAAACCGACTCCAGAACCTAGTCCAAGT 1184  
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374  
 Db 1185 CCGCAACTGTCACCAAAATCTCAACAGCTCCAAAGCAATCAATATGAT-----GAGAAA 1238  
 QY 375 LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394  
 Db 1239 TTGTCACAAAGAGCTGTTCGAAGAGTAGGCGATGTTAIGTCTTTGAGGAAATGGAGTT 1298  
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414  
 Db 1299 TCTCGTTATATCCAGCGCAAGAACTCTTTCAGCAGAAACAGCAGCAGGATTTGATAGCANA 1358  
 QY 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434  
 Db 1359 CTGCGCAAGCAGGAAAGTATTATCTCAAGCTAGGAGCTAAGAAACCTGACCTCCCATCT 1418  
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeu 454  
 Db 1419 AGTGATCGAGAATTTTACAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTA 1478  
 QY 455 Phe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474  
 Db 1479 CTTGATAATAAAGGTCACAAAGTTGATTTTCAGGCTTTGGATAACCTGTTGGACGACTC 1538  
 QY 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAlaPro 494  
 Db 1539 AAGATGCTCAAGTGTATAAGTCAAGTGTAGTGATGATATCTTTCCTTCTTAGCTCCG 1598  
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514  
 Db 1599 ATTCTGTCATCCAGACGTTTAGGAAACCAAAATGCGCAAAATACCTACACTGATGATGAG 1658  
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534  
 Db 1659 ATTCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACAGACGCTTATATCTTTGATCCT 1718  
 QY 535 HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554  
 Db 1719 CGTGATATAACAGTATGAGGGGATGCTATGTAACTCCACATATGACCCATAGCCAC 1778  
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys 574  
 Db 1779 TGGATTAAAAAGATAGTTGCTGAGCTGAGAGCGCGCAGCGCTTATGCTAATAA 1838  
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594  
 Db 1839 GAGAAAGGTTTACCCCTCTCCACAGACCATCAGGATTCAGGAATATCTGAGCAAAA 1898  
 QY 595 SerAlaAlaAlaTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu 614  
 Db 1899 GGAGCAGAGCTATCTACAAACCGCGTAAAGCAGCAGTAAAGAGGTGCGCACTGATCGATG 1958  
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634  
 Db 1959 CTTTACAATCTTCAATATCTAGTAGAGTCAAAAACGGTAGTTTAAATCATACCTCATAT 2018  
 QY 635 AspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsn 654

Db 2019 GACCATTACCAATAACATCAAAATTTGAGTGGTTGAGGAAGGCTTTATGAGGCACCTAAG 2078  
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674  
 Db 2079 GGTATATCTCTTGAGATCTTTTGGCGAGTGTCAAGTACTATGTGAAACATCCAAACGAA 2138  
 QY 675 ArgProHisSerAsnAspGlyTyrPglyAsnAlaSerGluHisValLeuGlyLysLysAsp 694  
 Db 2139 CGTCGCACTTCAGATAATGTTTGGTAACGCTAGCAGCATGTTCAAGAAGAACAAATAAT 2198  
 QY 695 HisSerGluAspProAsnLysAsnPhelysAlaAsp----- 706  
 Db 2199 GGTCAAGCTGATACCAATCAAAACGAAAAACCAAGCAGGAGAGAACTTCAGACAGAAAA 2258  
 QY 706 ----- 706  
 Db 2259 CCTGAGGAGAAACCCCTCGAGAGAGAGAAACCAAAACGAGAACCCAGAGTCTCCAAAA 2318  
 QY 707 -----GluGluProValGluGluThrProAlaGluProGluValProGluValGluThr 724  
 Db 2319 CCAACAGAGGAAACCAAGAGAAATCACCAGAGGAATCAGAGAACCTCAGTCGAGACT 2378  
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 Db 2379 GAAAGGTTGAAGAAAACTCAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2438  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGln 764  
 Db 2439 ATTATCAAGTCCATGCAAGAGACTCTCAGAGATTAAAAATAATTACTATTGGC 2498  
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 Db 2499 ACCAGGACCAACAATACTATTATGGCAGAGCTGAAAAACTATTGGCTTTTAAAGAG 2558  
 QY 785 Ser 785  
 Db 2559 AGT 2561  
 RESULT 19  
 BD229972  
 LOCUS  
 DEFINITION Human complement C3-degrading polypeptides from streptococcus pneumoniae.  
 ACCESSION BD229972  
 VERSION BD229972.1 GI:33039742  
 KEYWORDS JP 2002526082-A/3  
 SOURCE Streptococcus pneumoniae  
 ORGANISM Streptococcus pneumoniae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 2478)  
 AUTHORS Hostetter,M.K., Finkel,D.J., Cheng,Q., Green,B.A. and Masi,A.W.  
 TITLE Human complement C3-degrading polypeptides from streptococcus  
 JOURNAL Patent: JP 2002526082-A 3 20-AUG-2002;  
 REGENTS OF THE UNIVERSITY OF MINNESOTA,AMERICAN CYANAMID CO  
 COMMENT OS Streptococcus pneumoniae  
 PN JP 2002526082-A/3  
 PD 20-AUG-2002  
 PF 24-SEP-1999 JP 2000574269  
 PR 24-SEP-1998 US 60/101736,31-MAR-1999 US 09/283094 PI  
 MARGARET K HOSTETTER, DAVID J FINKEL, QI CHENG, BRUCE A GREEN, AMY  
 PI W MASI  
 PC C12N15/09,A61K38/00,A61K39/09,A61K39/395,A61K39/395,A61K48/00,  
 PC A61P29/00,  
 PC A61P37/04,A61P37/06,A61P43/00,C07K16/40,C12N1/15,C12N1/19, PC  
 C12N1/21,  
 PC C12N5/10,C12N9/48,C12P21/08,C12N15/00,C12N5/00,A61K37/02 CC  
 Human complement C3-degrading polypeptides from streptococcus CC  
 pneumoniae  
 FH Key Location/Qualifiers  
 FT source l. .2478  
 FT /organism='Streptococcus pneumoniae'.

Fri Oct 1 15:35:51 2004

# FEATURES

source Location/Qualifiers  
1. .2478  
/organism="Streptococcus pneumoniae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1313"

## ORIGIN

Alignment Scores: 2.35e-152 Length: 2478  
Pred. No.: 2812.00 Matches: 552  
Score: 77.10% Conservative: 81  
Percent Similarity: 77.10% Mismatches: 136  
Best Local Similarity: 67.24% Indels: 52  
Query Match: 67.52% Gaps: 6  
DB:

US-09-765-271-56 (1-796) x BD229972 (1-2478)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
Db 61 TCCTATGAACCTGGTGGTCCCAAGCTGGTCCAGGTTAAGAAGAGTCTAATCGAGTTTCT 120  
Qy 20 TyrlleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
Db 121 TATATAGATGGTATCAGCTGGTCAAAAGGCAGAAACTTGACACCAAGATGAAGTCAGT 180  
Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
Db 181 AAGAGGAGGGGATCAAGCCGCAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG 240  
Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
Db 241 ACCTCTCATGGAGACCAATATCATTAATATGCAAGGTCCTTATGATGCCATCATC 300  
Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
Db 301 AGTGAAGAGCTCTCATGAAGATCCGAATTCATGTTCAAGGATTCAGACATTCGTCAAT 360  
Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119  
Db 361 GAAATCAAGGGTGGTTATCTTATCAAGGTAGATGGAAATATCTATGTTTACCTTAAGGAT 420  
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
Db 421 GCAGCTCATCGCGATTAATTCGGACAAAGAGAGATTAACGTCAGAGCAGGAACAC 480  
Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
Db 481 AGTCATAATCACGGGGTGGTTCT---AACGATCAAGCAGTAGTTGCGAGCCAGAGCC 534  
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
Db 535 CAAGGACGGCTATACACGGATGATGTTATCTTCAATGCATCTCATATCATTTAGGAC 594  
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
Db 595 ACGGGTGATGCTTATATTCGTTCTCTACGGCAGCACATACATTCATCTTAAGATGAG 654  
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
Db 655 TTATCAGCTAGCAGTTAGTGTGTCGAGAAGCCTATTGGAATGGG----- 699  
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
Db 700 -----AAGCAGGGATCTCGTCTTCTTCAAGTCTTAGTTATATGCA 741  
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
Db 742 AATCCAGCTCAACCAAGATTGTCAGAGNACCAATCTGATGTCATCTCAACTTATCAT 801  
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
Db 802 CAA---AATCAAGGGGAAACATTTCAAGCCCTTTTACGTGAATTTGATGCTAAACCCCTTA 858

Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
Db 859 TCAGAACGCCATCTGGAATCTGATGGCTTATTTCCAGCCAGCGCAATACCAAGTCGA 918  
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
Db 919 ACCGCCAGAGGTGATGCTGCTCTCATGGTAACCATACCATTTATCCCTTATGAACA 978  
Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
Db 979 ATGTCGAATGGAAACAGATTTGCTCGTATTATTCCTCTCGTTATCGTTCAACCAT 1038  
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
Db 1039 TGGGTACAGATTCAAGACCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTTAGTCCA 1098  
Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
Db 1099 AGTCCGCAACCTGCACCAATCTCAACAGCTCCACCAATCCCAATTTGAT-----GAG 1152  
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
Db 1153 AAATTTGGTCAAGAGAGCTGTTTCGAAAGTAGTGGCATGGTTATGCTTTTGGAGAGAATGGA 1212  
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
Db 1213 GTTCTCTGTTATATCCCAAGCAAGATCTTTTCAGAGAAACAGCAGCAGGAGCATTTGATAGC 1272  
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
Db 1273 AAATGGCCAGCAGAGAAAGTTTATCTCATAGCTTAGGAGCTAAGAAACAGACCTCCCA 1332  
Qy 434 ProArgAspGlnGluPheTyrAsnLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
Db 1333 TCTAGTGTATCGAGATTTTACAATTAAGCTTATGACTTACTAGCAAGAATTCACCAAGAT 1392  
Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
Db 1393 TTACTTTGATTAATAAAGTTCGACAAAGTTTATGAGCTTTTGGATTAACCTGTTGGACGA 1452  
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAla 493  
Db 1453 CTCAGAGATGTCCTCAAGATGAATTAAGTCAAGTTAGTGGATGATATCTTCTGCTTCTTAGCT 1512  
Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
Db 1513 CCGATTCGTCATCCAGAACGTTTAGMAAACCAAAATGCGCAAAATACCTACACTGATGAT 1572  
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
Db 1573 GAGATTCAGTAGTCCCAAGTTGGCAGCAAGTACACCAACAGACAGCGTTATATCTTTGAT 1632  
Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
Db 1633 CCTCGTGAATTAACAGTATGAGGGGATGCTATGTAACCTCCACATATGACCCATAGC 1692  
Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
Db 1693 CACTGGATTAATAAAGATAGTTTCTGTAAGCTGAGAGAGCGCAGCCAGGCTTATGCT 1752  
Qy 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593  
Db 1753 AAAGAGAAAGGTTTGAACCTCTCTGACAGACCATCAGGATTCAGGAATACTAGGCA 1812  
Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613  
Db 1813 AAAGGAGCAGAACTATCTACACCGGCTGAAAGCAGCTAAGAAGGTGCCACTTGAATCGT 1872  
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
Db 1873 ATGCTTCAATCTTCAATATCTGTAAGAGTCAAAACCGGTAGTTAATCATACCTCAT 1932  
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653



```

Db      1933  TATGACCATTTACCAATCAATTTGAGTGGTTTGACGAAGCCCTTATGAGGACCT 1992
QY      654  AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db      1993  AAGGGGTATACTCTTGAGGATCTTTGGCGACTGTCAAGTACTATGCGAACATCCAAAC 2052
QY      674  GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693
Db      2053  GAACGTCGCGCATTCAGATAATGTTTGTGTAACGCTAGCGACCATGTTCAAGAACAACA 2112
QY      694  AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu 708
Db      2113  AATGGTCAAGCTGATACCAATCAACGGAACCAAGCGAGAGAAACCTCAGACAGAA 2172
QY      709  ---ProValGluGluThr 713
Db      2173  AAACCTGAGGAAGAACCCTCGAGNAGAGAAACCGCAAGGAGAGAAACCGAGTCTCCA 2232
QY      714  -----ProAlaGluProGluValProGluValGluThr 724
Db      2233  AAACCAACAGAGAAACCAAGAAATCACCAGAGGAATCAGAGAAACCTCAGTCCGAGCT 2292
QY      725  GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
Db      2293  GAAAGGTTTGAGAAAACTGAGAGAGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2352
QY      745  SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuAlaGlyLeuAlaGlyLeuGln 764
Db      2353  ATTATCAAGTCCATGCCAAGAGACTCTCAGAGATTAAAAATAATTTACTATTTGGC 2412
QY      765  IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly 784
Db      2413  ACCCAGACAACTACTATTATGCCAGAACTGAAAACTATTGGCTTTATTAAAGAG 2472
QY      785  Ser 785
Db      2473  AGT 2475

RESULT 20
AX569137
LOCUS      2457 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 2345 from Patent WO02077021.
ACCESSION AX569137
VERSION    AX569137.1 GI:26002635
KEYWORDS
SOURCE     Streptococcus pneumoniae
ORGANISM   Streptococcus pneumoniae
REFERENCE  1
AUTHORS    Masignani, V., Tettelin, H. and Fraser, C.
TITLE      Streptococcus pneumoniae proteins and nucleic acids
JOURNAL    Patent: WO 02077021-A 2345 03-OCT-2002;
          Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES   Location/Qualifiers
            source
              1..2457
              /organism="Streptococcus pneumoniae"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1313"
ORIGIN
Alignment Scores:
Pred. No.: 1.58e-151 Length: 2457
Score: 2797.50 Matches: 546
Percent Similarity: 77.30% Conservative: 84
Best Local Similarity: 66.92% Mismatches: 138
Query Match: 67.17% Indels: 47
DB: 6 Gaps: 7

US-09-765-271-56 (1-796) x AX569137 (1-2457)
QY      1  SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19

```

```

Db      61  TCCTATGAGCTTGAGCGTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120
QY      20  TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db      121  TATATAGATGGTGATCAGGCTGGTCAAAAGGCGAGAAACTTGACACAGATGAAGTCAGT 180
QY      40  LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db      181  AAGAGGAGGGGATCAACGCCCAACAAATGTTATCAAGATTACGGATCAAGGTTATGTG 240
QY      60  ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeu 79
Db      241  ACCTCTCATGGAGACCAATTATCTATATAATGGCAAGGTTCTTATGATGCCATCATC 300
QY      80  SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db      301  AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGGATTTCAGACATTGTCAAT 360
QY      100  GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db      361  GAAATCAAGGGTGGTTATGTCTATTAAGGTAACGGTAATACTATGTATTACCTTAAGGAT 420
QY      120  AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db      421  CGAGCTCATGCGGATAATATTCGGACAAAAGAGAGATTAAACGTCAGACAGAGAAACGC 480
QY      140  SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db      481  AGTCATTAATCATACTCA-----AGAGCAGATAATGCTGTGTGTCAGCCAGAGACC 531
QY      160  GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db      532  CAAGGAGCTTATACACGAGATGAGGTATATCTTCAATGCATCTGATATCATTCATTGAGAC 591
QY      180  ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db      592  ACGGCTGATCTTATATCGTTCTCTCACGGCGACCAATTACCATACATTCTTAAGAAATGAG 651
QY      200  LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db      652  TTATCAGCTAGCGAGTTAGTCTGTCAGAAAGCCTATTGGAATGGG-----696
QY      220  AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db      697  -----AAGCAGGAGATCTCGCTCTCTTCAAGTTCTTCAAGTTCTTAATAATGCA 738
QY      240  SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db      739  AATCCAGCTCAACCAAGATTGTACAGAGAACCAATCTGACTGTCTCACTCAACTTATCAT 798
QY      260  GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db      799  CAA---AATCAAGGGGAAACCAATTTCAGCCCTTTTACGTGAATTGTATGTAACCCCTTA 855
QY      280  SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db      856  TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCTGA 915
QY      300  ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db      916  ACCGCCAGAGGTGAGTGTCCCTCATGGTAACCAATTACCACTTTATCCCTTATGACAA 975
QY      320  MetSerGluLeuGluGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339
Db      976  ATGTCGTAATTGGAAAAACGAATTCGCTGATTATTCCTCTTCGTTATCGTTCAAAACCAT 1035
QY      340  TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db      1036  TGGGTACCAGATTCAAGACCAAGAGAACCAAGTCCACAGACCTCCAGAACCTAGTCCA 1095
QY      360  GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379

```

1096 AGTCGCAACAGCTCCAGCAATCAATGAT-----GAGAAATGGTCAAGAAGCT 1149  
 Db  
 380 ValArgLysValGluGluGlyTyrValPheGluGluGlySerArgTyrValPhe 399  
 QY  
 1150 GTTCGAAAGTAGCGATGTTATGCTTTGAGGAGATGGAGTCTCTGTTATATCCCA 1209  
 Db  
 400 AlAlAspLeuProSerGluThrValLysAsnLeuGluSerLysLysGluGlu 419  
 QY  
 1210 GCCAAGATCTTCAGCAGAAACAGCAGCAGGATTCATAGCAATGGCCAGCAGAA 1269  
 Db  
 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgGluGluPhe 439  
 QY  
 1270 AGTTATCTCATAGCTAGGAACCTAAGAACTGACCTCCCATCTAGTATGTCGAGATT 1329  
 Db  
 440 TyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 QY  
 1330 TACAATAGGCTTATGACTTACTAGCAAGATTCCACCAAGATTTTATGATAATAAAGGT 1389  
 Db  
 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThr 479  
 QY  
 1390 CGACAAGTTGATTTAGGCTTTGATTAACCTGTTGGAACGACTCAAGGATGCTCAAGT 1449  
 Db  
 480 AsnLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 QY  
 1450 GATAAAGTCAAGTTAGTGAAGATATCTTGCTTCTTAGCTCCGATTCGTCATCCAGAA 1509  
 Db  
 500 ArgLeuGlyLysProAsnSerGlnIleGluThrGluAspGluValArgIleAlaGln 519  
 QY  
 1510 GTTTAGGAAACCAATGCGCAATTACTTACCTGATGATGATGATCAAGTAGCCAAAG 1569  
 Db  
 520 LeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSer 539  
 QY  
 1570 TTGGCAGGCAAGTACACACAGAACGCGTTATCTTTGATCCTCGTATATACCAAGT 1629  
 Db  
 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559  
 QY  
 1630 GATGAGGGGGATGCTTATGATCTCCATATATGACCATGACCATGATTAATAAAGAT 1689  
 Db  
 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579  
 QY  
 1690 AGTTTGTCTGAGCTGAGAGCGCGCAGCCAGGCTTATGCTTAAGAGAAAGGTTTGACC 1749  
 Db  
 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599  
 QY  
 1750 CTTCTCTTCAGACACCATCAGGATTCAGGAATCTGAGCAGAAAGAGCAGAGCTATC 1809  
 Db  
 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619  
 QY  
 1810 TACAACCGGTGAAGACGCTAAGAGGTGCCATTCATGCTATGCTTACATCTTCAA 1869  
 Db  
 620 HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn 639  
 QY  
 1870 TATACCTGTAGAGTCAAAACCGTAGTTTAAATCATACCTCATGATGATCATTACCAAT 1929  
 Db  
 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659  
 QY  
 1930 ATCAAAATTTGAGTGGTTTGAAGAGGCTTTATGAGGCACCTAAGGGGTATATCTCTGAG 1989  
 Db  
 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679  
 QY  
 1990 GATCTTTTGGCGACTGTCAGTACTATGTGCAACATCCAAACGAACTGCGCATTCAGAT 2049  
 Db  
 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspPro 699  
 QY  
 2050 AATGGTTTGTGTAACGCTAGCAGCATGTTCAAGAGAAACAAAAATGGTCAAGCTGATACC 2109  
 Db  
 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713  
 QY  
 2110 AATCAACCGGAAACCAAGCGAGAGAAACCTGACAGAGAAACCTGAGAGAAAC 2169  
 Db  
 713 ----- 713  
 2170 CCTCGAGAGAGAAACCGCAAGCGAGAAACAGAGATCTCCAAACCAACCAAGAGAACCA 2229  
 Db

714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730  
 QY  
 2230 GAAGATCACCAGGAGAACTCAGAGAACTCAGTCCAGACTGAAAGGTTTGAAGAAAA 2289  
 Db  
 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750  
 QY  
 2290 CTGAGAGAGGCTGAAGATTCTTGGAAATCCAGATCCAGATCCATATATCAAGTCCAATGCC 2349  
 Db  
 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770  
 QY  
 2350 AAGAGAGCTCTCAGAGATTAAAAATAATTTACTATTGGCACCAGGACACAAATACT 2409  
 Db  
 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785  
 QY  
 2410 ATTATGCGAGAGCTGAAATACTATTGGCTTTATTAAGGAGAGT 2454  
 Db  
 RESULT 21  
 BD264091 2481 bp DNA linear PAT 17-JUL-2003  
 LOCUS Streptococcus pneumoniae proteins and nucleic acid molecules.  
 DEFINITION BD264091  
 ACCESSION BD264091  
 VERSION BD264091.1 GI:33073859  
 KEYWORDS JP 2002531055-A/27.  
 SOURCE Streptococcus pneumoniae  
 ORGANISM Streptococcus pneumoniae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 2481)  
 AUTHORS Gilbert,C.F.G. and Hansbro,P.M.  
 TITLE Streptococcus pneumoniae proteins and nucleic acid molecules  
 JOURNAL Patent: JP 2002531055-A 27 24-SEP-2002;  
 MICROBIAL TECHNIQS LTD  
 COMMENT OS Streptococcus pneumoniae  
 PN JP 2002531055-A/27  
 PD 24-SEP-2002  
 PF 27-JUL-1999 JP 2000562519  
 PR 27-JUL-1998 GB 9816337.1,19-MAR-1999 US 60/125164 PI  
 CHRISTOPHE FRANCOIS GUY GILBERT,PHILIP MICHAEL HANSBRO PC  
 C12N15/09,C12N15/09,A61K39/09,A61K45/00,A61K48/00,A61P31/04, PC  
 C07K14/315.  
 PC C07K16/12,C12Q1/14,C12Q1/68,G01N33/53,G01N33/566,G01N33/569//  
 PC C12P21/06,C12Q1/15,G01N33/50,(C12N15/09,C12R1:46),(C12Q1/14,C12R1:46),  
 PC G01N33/15,G01N33/50, PC  
 C12N15/00, PC  
 C12N15/00,(C12N15/00,C12R1:46)  
 CC Streptococcus pneumoniae proteins and nucleic acid molecules  
 FH key Location/Qualifiers  
 FT source 1..2481  
 FT /organism='Streptococcus pneumoniae'.  
 FEATURES  
 source  
 1..2481  
 /organism='Streptococcus pneumoniae'  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:1313'  
 ORIGIN  
 Alignment Scores: 2,22e-151 Length: 2481  
 Pred. No.: 2795.00 Matches: 549  
 Percent Similarity: 76.98% Conservative: 83  
 Best Local Similarity: 66.87% Mismatches: 137  
 Query Match: 67.11% Indels: 52  
 DB: 6 Gaps: 8  
 US-09-765-271-56 (1-796) x BD264091 (1-2481)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19  
 Db 61 TCCTATGAGCTTGTGCTGCTCAACAGCTGGTCAAGATGAGATGAGATGCTATCGAGTTGCT 120  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39



121 TATATAGATGGTGCATCAGGCTGGTCAAAAGGCGAGAAAACCTTGACACCGAGTGAAGTCAGT 180  
40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
181 AAGAGGGAGGGGATCAACCGCCGACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG 240  
60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
241 ACCCTCATGGAGACCAATATCATTTACTATATATGCAAGGTCCTTATGATGCCATCATC 300  
80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
301 AGTGAAGAGCTCCTCATGAAGATCCGAATATCATAGTTGAAGGATTCACATTTGTCAAT 360  
100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
361 GAAATCAAGGGTGGTTATGTCTATCAAGGTAGACGGAAATATCATGTTTACCTTAAGGAT 420  
120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
421 GCAGCTCATGCGGATAATATTCGACAAAAGAGAGATTAAAGCTCAGAAGCAGCAAGCC 480  
140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
481 AGTCATATATCAC-----GGGTCAAGAGCTAACCATCATGTCAGTAGCTGCAGCCAGAGCC 534  
160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
535 CAAGGACGCTATACAACGGATGATGGGTATATCTTCAATGTCATCTGATATCATTTAGGAC 594  
180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
595 ACGGGTCATGCTTATATCGTTCCTCACGGCAGCATTAACCATTAACATTCCTAAGAAATGAG 654  
200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
655 TTATCAGTACGAGTGTAGCTGCTGCGAAGCCCTATTGGAATGGG----- 699  
220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
700 -----AAGCAGGATCTCGTCTCTTCTCAAGTCTCTAGTTCTAATGCA 741  
240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
742 AATCCAGCTCAACCAAGATGTGCAGAACCAACAATCTGACTGTCACTCCCACTTATCAT 801  
260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
802 CAA--AATCAAGGGGAACAACTTCAAGCCCTTTTACGTGAATGTATGCTAAACCCCTTA 858  
280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
859 TCAGAACGCCCATGTGAAATCTGATGCGCTTATTTTCGACCCAGCGCAATTCAGTCCGA 918  
300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
919 ACCGCCAGAGGTAGCTGCTCCCTCATGTGTAAACCATTAACACTTTATCCCTTATGAACA 978  
320 MetSerGluLeuGluGluAlaArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
979 ATGCTGTAATTGGAAAGAAAGAAATGTCTGTAATTTATTCCTCTGTTATGCTTCAAAACCAT 1038  
340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
1039 TGGGTACCAAGATTCAAGACCAAGACCAACCAAGTCCCAATTCGACTCCGGAACCTAGTCCA 1098  
360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
1099 AGTCGCGCAACCTGCGCAAAATCTCAACCGAGCTCCAAAGCAATCCAAATGAT-----GAG 1152  
374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGlyLysGly 393  
1153 AAATGGTCAAGAAAGCTGTTGCAAAAGTAGGCGATGTGTCTTTGAGGAGAATGCA 1212

394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
1213 GTTCTCTCGTTATCCAGCCCAAGGATCTTTTTCAGCAGAAACAGCAGCAGCGATTGATAGC 1272  
414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
1273 AAACCTGGCCACAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAAACTGACCTCCCA 1332  
434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
1333 TCTAGTGTATCGAAGATTTTACAATAAGGCTTATGATCTTACTTACGAAGAATTCACCAAGT 1392  
454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
1393 TTACTTGTATAATAAAGGTGCAAGTTGATTTTGGGGTTTGGATAACTGTTTGAACGA 1452  
474 LeuAsnAspGluSerThrAsnLysGlnLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
1453 CTCAAGGATGTCCCAAGTGTATAAAGTCAAGTTAGTGGATGATATTTCTTGCTTCTTAGCT 1512  
494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
1513 CCGATTCGTCTATCCAGAACGTTTAGGAAAACCAATGCGCAATTTACCTACACTGATGAT 1572  
514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
1573 GAGATTTCAAGTAGGCAAGTTTGGCAGGCAAGTACACACAGAACGCGTTATATCTTTGAT 1632  
534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
1633 CCTCGTGTATATAACCAAGTATGAGGGGATGCTTATGTAATCCACATATATGACCATAGC 1692  
554 HistripIleGlyLysAspSerLeuSerAspLysGlnLysValAlaAlaGlnAlaTyrThr 573  
1693 CACTGGATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGAGCCAGGCTTATGCT 1752  
574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
1753 AAAGAGAAAGGTTTGAACCTCTCTCGACAGACCATCAGGATTCAGGAATATCTAGGCA 1812  
594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613  
1813 AAGAGGACGAAAGCTATCTACAAACCGCTGAAAGCAGCTAAGAAAGGTGCACCTTATCGT 1872  
614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
1873 ATGCTTACAATCTCAATATCTGTAGAACTCAAAACCGTAAACCGTAACTATATCATACCTCAT 1932  
634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
1933 TATGACCATTTACATAACATCAAAATTTGAGTGGTTTGCAGAGGCGCTTTTATGAGGACCT 1992  
654 AsnGlyTyrThrLeuGluAspPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
1993 AAGGGGTATACTCTTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 2052  
674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
2053 GAACGTCGCGCATTCAGATAATGTTTGTAAACGCTAGCAGCATGTTTCAAAAGAAACAAA 2112  
694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708  
2113 AATGTCAGCTGTATACCAATCAAAACGGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAA 2172  
709 ---ProValGluGluThr----- 713  
2173 AAACCTGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAAAGCGAGAAACCGAGGTCTCCA 2232  
714 -----ProAlaGluProGluValProGlnValGluThr 724  
2233 AAACCAACAGAGGAACCAAGAAATTCACGAGGAATCAGAAAGAACCTCAGGTCGAGACT 2292



QY 160 GlnGlyArgThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuGluAsp 179  
 Db 532 CAAGGAGCTTTATACAAAGGATGATGGTATATCTTCAATGCATCTCATATCATTTGAGGAC 591  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProIysAsnGlu 199  
 Db 592 ACGGTGATGCTTATATCGTCTCCTCAGCGGACCATTTACATTTACATTTCTTAAGAAATGAG 651  
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 652 TTATCAGCTAGCGAGTTAGCTGCTCAGAAAGCTATTTGGAATGGG----- 696  
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 Db 697 -----AAGCAGGATCTCTGCTCTTCAAGTTCTAGTTTATATGCA 738  
 QY 240 SerValSerAspProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCACTCAACTTATCAT 798  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db 799 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATGCTTAAACCCCTTA 855  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db 856 TCAAGACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCGCAATCACAGTCA 915  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db 916 ACCGCCAGAGGTGAGCTGCTCCCTCATGGTGAACCATTTACCACTTTATCCCTTATGAACA 975  
 QY 320 MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 Db 976 ATGCTCAATTGGAAAAACGAATTTGCTGTATTTCCCTTCGTGTATCGTTTCAAAACCAT 1035  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1036 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCACACCGACTCCAGAACCTTAGTCCA 1095  
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerLeuValSerGlnLeu 379  
 Db 1096 AGTCCGCAACCACTCCAGCAATCCAAATTGAT-----GGGAATTTGGTCAAAAGACT 1149  
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399  
 Db 1150 GTTCGAAAAGTAGCGATGCTTATGCTTTGAGAGAGATGAGTTTCTCGTTATATCCCA 1209  
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
 Db 1210 GCCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCAATTTAGCAAACTGGCCCAAGCAGAA 1269  
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439  
 Db 1270 AGTTTATCTCATAGTAGGAATAGAAACCTAGAACTCCCACTCCCACTAGTATCAGAAATTT 1329  
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 Db 1330 TACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGAATAAAGGT 1389  
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
 Db 1390 CCACAAGTTGATTTTGGGCTTTGGATAACCTTGTGGAACGACTCAAGATGTCCTCACT 1449  
 QY 480 AsnLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 Db 1450 GATAAAGTCAAGTTAGTGAAGATATTTGCTTCTTAGCTCCGATTCGTATCCAGAA 1509  
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519  
 Db 1510 CGTTTAGGAAAAACCAATCGCAATTTACCTACATGATGAGATCAAGTAGCCCAAG 1569  
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539

Db 1570 TTGGCAGGCAAGTACACAGCAGAACGGTATATCTTTGATCCTCGTGATATAACCACT 1629  
 QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559  
 Db 1630 GATGAGGGGATGCTTATGTAATCTCCACATATGACCATAGCCACTGGAATTTAAAAAGAT 1689  
 QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579  
 Db 1690 AGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGCTTATGCTTAAGAGAAAGTTTGACC 1749  
 QY 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599  
 Db 1750 CTTCTTCGACAGACCATCAGGATTCAGAGAAATCTGAGGCAAAAGGAGCAAGAGTATC 1809  
 QY 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619  
 Db 1810 TACAACCGGTGAAAGCAGCTTAAGAAAGTGCACCTTGATCGTATGCTTACAAATCTCAA 1869  
 QY 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639  
 Db 1870 TATACCTGTAGAAGTCAAAACCGTAGTTTATCATCTCATCTATGACCATTTACAAATCTCAA 1929  
 QY 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659  
 Db 1930 ATCAAAATTTGAGTGGTTTGAAGAGGCTTTATGAGGCACCTAAGGGGTATACCTTTGAG 1989  
 QY 660 AspLeuPheAlaTrpIleLysTyrValGluHisProAspGluArgProHisSerAsn 679  
 Db 1990 GATCTTTGGGCACTGTCAAGTACTATGTGGAACATCCAAACGACGCTCCGATTCAGAT 2049  
 QY 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699  
 Db 2050 AATGGTTTGGTAACGCTAGCGACCATGTTCAAGAAACAAATAAGTGGTCAAGCTGATACC 2109  
 QY 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713  
 Db 2110 AATCAAAACGGAAAAACCAAGCGAGGAAACCTCAGACAGAAAAACCTGAGGAGAAACCC 2169  
 QY 713 ----- 713  
 Db 2170 CCTCGAGAGAGAAACCGCAAGCGAGAAACCAAGAGTCTCCAAACCAACAGAGAAACCA 2229  
 QY 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730  
 Db 2230 GAAGAATCACAGAGAAATCAGAGAAACCTCAGTGCAGACTGAAAAGGTGAAGAAAAA 2289  
 QY 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750  
 Db 2290 CTGAGAGGCTGAAGATTTACTTTGAAAAATCCAGATCCCAATTTCAAGTCCAATGCC 2349  
 QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770  
 Db 2350 AAAGAGACTCTCAGAGATTAATAATAATTTACTATTGTCACCCAGCAGCAACAATACT 2409  
 QY 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785  
 Db 2410 ATTATGGCAGAGCTGAAAAAACTATTGGCTTTATTAAAGGAGAGT 2454

## RESULT 23

BD268049

LOCUS

DEFINITION

2531 bp DNA linear

PAT 17-JUL-2003

Streptococcus pneumoniae protein and immunogenic fragments for

vaccines.

ACCESSION

BD268049

VERSION

BD268049.1

GI:33077817

KEYWORDS

JP 2002532561-A/7.

SOURCE

Streptococcus pneumoniae

ORGANISM

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

REFERENCE

1 (bases 1 to 2531)

AUTHORS

Johnson, L.S., Koenig, S. and Adamou, J.E.

Streptococcus pneumoniae protein and immunogenic fragments for  
 Patient: JP 2002532561-A 7 02-OCT-2002;  
 MEDIMUNE INC  
 COMMENT  
 OS Streptococcus pneumoniae  
 PN JP 2002532561-A/7  
 PD 02-OCT-2002  
 PF 21-DEC-1999 JP 2000589215  
 PR 21-DEC-1998 US 60/113048  
 PI LESLIE S JOHNSON, SCOTT KOENIG, JOHN E ADAMOU  
 PC A61K39/09, A61K38/00, A61P31/10, A61K37/02  
 CC Streptococcus pneumoniae protein and immunogenic fragments for

CC Key vaccines  
 FH Location/Qualifiers  
 FT source 1. .2531  
 FT /organism='Streptococcus pneumoniae'  
 Location/Qualifiers  
 1. .2531  
 /organism='Streptococcus pneumoniae'  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:1313'

ORIGIN

Alignment Scores:  
 Pred. No.: 6,99e-151 Length: 2531  
 Score: 2786.50 Matches: 544  
 Percent Similarity: 77.06% Conservative: 84  
 Best Local Similarity: 66.75% Mismatches: 140  
 Query Match: 66.90% Indels: 47  
 DB: 6 Gaps: 7

US-09-765-271-56 (1-796) x BD268049 (1-2531)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19  
 DB 61 TCCTATGAGCTTGGAGCTTACCAAGCTGCTCAGGATAGAAAGAGTCTAATCGAGTTGCT 120  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 DB 121 TATATAGATGGTGATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACCATGAGTCACT 180  
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 DB 181 AAGAGCGGGGATCAACGCCGAACTTATTCAGATTAACGATTAACGATCAAGTTATGG 240  
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 DB 241 ACCTCTCATGGAGACCATATCATCTACTATATATGGCAAGTTCCTTTATGATGCAATC 300  
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 DB 301 AGTGAAGAGCTCTCATGAAGATCCGAATATATGAGTTGAAGATTCAGACATTCAT 360  
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119  
 DB 361 GAAATCAAGGTTGGTATGTCATTAAGTAAACGTTAAATACTATGTTTACCTTAAGAT 420  
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 DB 421 GCGGCTCATGCGGATATATTCGCAAAAGAGAGATTAAACGTCAGAGCGAGGACGC 480  
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 DB 481 AGTCATAATCAATCACTCA-----AGAGCAGATATGCTGTGTGTCAGCGAGGCC 531  
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 DB 532 CAAGGAGCTTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGAC 591  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 DB 592 ACGGGTATGCTTATATCGTTCTCTCAGCGGACCATTAACCATTAATCAATTCCTAAGATGAG 651

200 LeuSerAlaSerGlnLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 DB 652 TTATCAGCTAGCGAGTTAGCTGCTCAGAGACCTATTTGGAATGGG----- 696  
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 DB 697 -----AAGCAGGGATCTCGTCTCTTCAAGTTCTAGTTATATGCA 738  
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 DB 739 AATCCAGCTCAACCAAGATTGTTCAGAGACCAACATCTGACTGTCTCACTCACTCACTATCAT 798  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 DB 799 CAA---AATCAAGGGGAAAAACATTTCAAGGCTTTTACGTAATTTGTATGCTAAACCTTCA 855  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 DB 856 TCAGAACGCCATGTGGAAATCTGATGGCTTATTTTCGACCCAGCGCAATACCAAGTCGA 915  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 DB 916 ACCGCCAGAGGTAGCTGCTCATGGTAACCATTAACCATTTATCCCTTATGAAACAA 975  
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 DB 976 ATGTCGAAATTCGAAAAACGAATTCGTCGTATTTATCCCTTCGTATTCGTTCAACCAT 1035  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 DB 1036 TGGGTACCAGATTCAAGACCGAGAACCAAGTCCACAAACGACTCCAGAACCTAGTCCA 1095  
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379  
 DB 1096 AGTCGCAACAGCTCCCAAGCAATCCAATTGAT-----GGGAAATTTGTCAAAAGCT 1149  
 QY 380 ValArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399  
 DB 1150 GTTCAAAAGTAGGCGATGTTATGCTTTGAGGAGAAATGGAGTTCTCTGTTATATCCCA 1209  
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLysSerLysLeuSerLysGlnGlu 419  
 DB 1210 GCCAAGGATCTTTCAGCAAAACAGCAGCAGCAGCATGATGATGATGATGATGATGATG 1269  
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439  
 DB 1270 AGTTTATCTCATAGCTAGGAACTAAGAAACCTGACCTCCCATCTAGTCTAGTCTAGT 1329  
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 DB 1330 TACAATAAGGCTTATGACTTACTAGCAAGATTCCCAAGATTACTTGTATATAAAGGT 1389  
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
 DB 1390 CGCAAGTTGATTTTGGAGCTTTGGATAACCTGTTCGACGACTCAAGGATGCTCAAGT 1449  
 QY 480 AsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 DB 1450 GATAAAGTCAAGTTAGTGAAGATATTTCTGCTCTTCTAGTCTGCTGCTGCTGCTGCTG 1509  
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519  
 DB 1510 CGTTTATGAGAAACCAATGCGCAATTTACCTACATGATGATGATGATGATGATGATG 1569  
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539  
 DB 1570 TTGGCAGCGCAATACACAGCAGAGACCGTTATATCTTTGATCTCTGCTGATATAACCAT 1629  
 QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAsp 559  
 DB 1630 GATGAGGGGATGCTTATGTAACCTCAATATGATGATGATGATGATGATGATGATGATG 1689  
 QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579



320	MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339	Db
976	ATGCTCAATTTGAAACAAAGATTCGCTGATTAATTCCTTCCTGTTTCAAAACCAT	1035	Db
340	TTPValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359	QY
1036	TGGGTACCATGATTCAGAACAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA	1095	Db
360	GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu	379	QY
1096	AGTCCGCAACAGCTCCAGCAATCCAAATGAT-----GGGAAATGGTCAAGAGCT	1149	Db
380	ValArgLysValGlyGluGlyTyrValPheGluGluGlyIleSerArgTyrValPhe	399	QY
1150	GTTTCAAAAGTAGGCGATGTTATGCTTTGAGAGAAATGGAGTTTCTGCTTATATCCCA	1209	Db
400	AlaLysAspLeuProSerGluThrValLysAsnLeuSerLysLysGlnGlu	419	QY
1210	GCCAAAGATCTTTCAGCAGAACACGACGAGCGATGATAGCAACATGGCCACAGCAGAA	1269	Db
420	SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe	439	QY
1270	AGTTTATCTCATAGCTAGGAACTAGAAACTGACCTCCCATCTAGTGTGAGAAATTT	1329	Db
440	TyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLysPhe***AsnLysGly	459	QY
1330	TACATAAGGCTTATGACTTACTAGCAAGAAATCCCAAGATTTTACTTGATAAAGGT	1389	Db
460	ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr	479	QY
1390	CGACAAAGTTGATTTTGAGGCTTTGGATAACTGTGTGAACGACTCAAGGATGTCACAGT	1449	Db
480	AsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu	499	QY
1450	GATAAAGTCAAGTAGTGAAGATATCTTGCCTCTTAGCTCCGATTCGTCATCCAGAA	1509	Db
500	ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln	519	QY
1510	CGTTTAGGAAACCAAAATGCGCAATATTACCTACACTGATGATGATTCAGATGACCAAG	1569	Db
520	LeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer	539	QY
1570	TTGGCAGGCAAGTACACAGCAGAACGGTATATCTTTGATCTCTGATATATACCAAGT	1629	Db
540	AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp	559	QY
1630	GATGAGGGGATGCCATGCTAATCTCCACATATGACCCATAGCCACTGGATTAAAAAGAT	1689	Db
560	SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu	579	QY
1690	AGTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCTTAAAGAAAGGTTGACC	1749	Db
580	ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle	599	QY
1750	CCTCCTTCAGACAGACCATCAGGATTCAGGAATATCTAGGCAAAAGGACGAGCTATC	1809	Db
600	TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu	619	QY
1810	TACAACCGMTGAACAGCTAAGAGGTCGCCATTCATGCTGATGCTTACCAATCTTCAA	1869	Db
620	HisThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsn	639	QY
1870	TATAGTCTAGAAAGTCAAAACGGTAGTTAATCATACCTCATATGACCATACCATAC	1929	Db
640	IleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu	659	QY
1930	ATCAAAATTTGAGTGGTTTGAAGAGCGCTTATAGGACCTTAAGGGGTATCTCTTGAG	1989	Db
660	AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn	679	QY
1990	GATCTTTTGGGACTGTCAAGTACTATGTGCAACATCCAAACGAAAGCTCGCATTCAGAT	2049	Db



## ORIGIN

## Alignment Scores:

Pred. No.: 4,18e-148 Length: 232807  
 Score: 2781.00 Matches: 543  
 Percent Similarity: 76.89% Conservative: 89  
 Best Local Similarity: 66.06% Mismatches: 136  
 Query Match: 66.77% Indels: 54  
 DB: 2 Gaps: 7

US-09-765-271-56 (1-796) x SPNEU1901 (1-232807)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19  
 Db TCCTATGAACCTGGCGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTATCGAGTTGCT 200412  
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 Db TATATAGATGGTATCATCGGCTGGTCAAAAGCGCAGAAAATTTGCACCAGATGAAGTCAGT 200472  
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db AAGAGGAGGGGATCAACGCCGAAACAAATGTTATCAAGATTACGGATCAAGGTTATGTG 200532  
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 Db ACCTCTCATGGAGACCATATATCTACTATAATGGCAAGGTTCTTATGATGCCATCATC 200592  
 Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGGATTCAGACATTTGTCAAT 200652  
 Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
 Db GAAATCAAGGGTGGTATGTCATTAAAGTAACGTAATTAATCTATGTTTACCTTAAAGGAT 200712  
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 Db GCAGCTCATCGGATATAATTTCGCAAAAAGAGAGATTAAACGTCAGAGCAGGAAACGC 200772  
 Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGTGTCAGCAGCAGAGCC 200823  
 Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db CAAGGACGTACACACAGATGATGTTATCTTTAAATGCTTCTGATATCATTTGAAGAT 200883  
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db ACTGGTATGCCCTATATCGTTCCTCATGGCAATCACTTCCACTATATTCCTAAGAGTGAT 200943  
 Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db TTATCTGCCAGTGAATAGTCTGCTGCCCAAGCCTATTGGAAATGGG-----200988  
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 Db -----AAGCAGGGATCTCGCTCTTCAAGTTCTAGTCAATAATGCA 201030  
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db AATCCAGCTCAACCAAGATTGTCAGAAACCAACAACTGACTGTCTACTCAACTTATCAT 201090  
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTAATTTGATGTAACCCCTTAA 201147  
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db TCAGAACGCCATGTGAATCTGATGCGCTTATTTTCGACCCAGCGCAATATCAAGTCGA 201207  
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319

Db 201208 ACCCCAGAGGTAGCTGCTCCCTCATGGTAACCATACCACCTTATCCCTTATGAACAA 201267  
 Qy 320 MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 Db ATGTCTCAATTTGAAACCAAGATGCTCGTATTATTCCTTCCCTTCGTTATCGTTCAACCAT 201327  
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db TGGGTACCAAGATTCAAGACCAGAACCAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 201387  
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 Db AGTCCGCAACTGCACCAAAATCTCAACAGCTCCCAAGCAATCCAAATGAT-----GAG 201441  
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGluLysGly 393  
 Db AATTTGGTCAAGAGCGTGTTCGAAAAGTAGGCGATGTTATGTCTTTGAGGAGAAATGGA 201501  
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuLeuSer 413  
 Db GTTTCCTCGTTATATCCAGCCAAAGGATCTTTCAGCAGAAAACAGCAGCAGCATTGATAGC 201561  
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 Db AACTGCCCAAGCAGGAAGTTTATCTCATAGCTAGAGCTAAGAAAACCTGACCTCCCA 201621  
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 Db TCTAGTGATCGAGATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 201681  
 Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArg 473  
 Db TTACTTCGATAATAAAGCGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGACGA 201741  
 Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
 Db CTCAGAGATGCTCAAGTGATAAAGTCAAGTTAGTGGAGATATTCTTGCCTTCTTAGCT 201801  
 Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
 Db CCGATTTCGTATCCAGAACGTTTAGCAAAACCAATTCGCAAAATTCCTACACTGATGAT 201861  
 Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
 Db GAGATTCAAGTAGCCCAAGTTGGCAGCAAGTACACAAACAGAACGCTTATATCTTTGAT 201921  
 Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 Db CCTCGTGATATAACCACTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 201981  
 Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
 Db CACTGGATTAAAAAAGATAGTTGCTGAAGCTGAAAGAGCGCAGCCCCAGGCTTATGCT 202041  
 Qy 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593  
 Db AAGAGAAAAGGTTTACCCTCTTCGACAGACCATCAGGATTTCAGGAAATACTAGGCA 202101  
 Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613  
 Db AAGAGGAGCAAGACTATCTACAAACCGCTGAAGCAGCTAAGAAAGTGCACCTTATGTCG 202161  
 Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
 Db ATGCCTTACAATCTCAATATACTGTAGAAGTCAAAAACGCTAGTTTATCATCATCTCAT 202221  
 Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653  
 Db TATGACCATTTACCAATAACATCAAAATTTGAGTGGTTTGACGAAGGCTTTTATGAGCACCT 202281  
 Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673



Db 202282 AAGGGTATACTCTTGAGGATCTTTGGCGACTGCTCAAGTACTATGTGACATCCAAAC 202341  
Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
202342 GAACGTCGGCATTGAGATAATGGTTTGGTAACGCTAGCGACCATGTTCAAAGAACAACAA 202401  
Qy 694 AspHisSerGluAspProAsnLysAsnPhelLysAlaAspGlu----- 707  
Db 202402 AATGGTCAAGCTGATACCAATCAAAACGGAACCAACGAGAGAGAAACCTCAGACAGAA 202461  
Qy 707 ----- 707  
Db 202462 AAACCTGAGGAGAAACCCCTCGAGAGAGAAACCGCAACGAGAAACAGAGTCTCCA 202521  
Qy 708 -----GluProValGluGluThrProAlaGluProGluValProGlnValGlu 723  
202522 AAACCAACAGACGAAACCAAGAGAGAAATCACAGAGGAATCAGAAGAACCTCAGGTCGAG 202581  
Qy 724 ThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743  
202582 ACTGAAAGGTTGAAGAAACTGAGAGAGGCTGAAGATTACTTTGGAAAAATCCAGGAT 202641  
Qy 744 SerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeu 763  
Db 202642 CCAATTATCAAGTCCCAATGCCAAAGAGACTCTCACAGGATTAAAAATAATTTACTATT 202701  
Qy 764 GlnIleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLys 783  
Db 202702 GGCACCCAGGACAACTACTATTATGTCAGAAAGCTGAAAACTATTGCTTTATTAAAG 202761  
Qy 784 GlySer 785  
Db 202762 GAGAGT 202767

Search completed: October 1, 2004, 10:19:24  
Job time : 8737 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 1, 2004, 07:35:01 ; Search time 748 Seconds  
(without alignments)  
4520.810 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 STELGLYQARTVKNRVS.....KLLALLKGNPSVSKEKIN 796

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool.h/US09765271/runat.30092004.113752.12270/app.query.fasta.1.967  
-DB=N\_Geneseq 29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS-human40.cdi  
-LIST=90 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25  
-MODE=LOCAL -OUTPT=ptp -NORM=ext -HAPSIZE=500 -MINLEN=0 -NCPUL=6 -ICPU=3  
-USER=US09765271 @CEN.1.1.513 @runat.30092004.113752.12270 -MAXLEN=2000000000  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq 29Jan04.\*

1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001as.\*  
5: Geneseq2001bs.\*  
6: Geneseq2002s.\*  
7: Geneseq2003as.\*  
8: Geneseq2003bs.\*  
9: Geneseq2003cs.\*  
10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	2389	2	Aav27351 Streptoco
2	4163	100.0	2389	6	Abq84819 S. pneumo
3	4163	100.0	2389	9	Adc45136 S. pneumo
4	4163	100.0	2406	7	Abx06886 S. pneumo
5	4163	100.0	2451	3	Aaa47604 Recombina
6	4163	100.0	8195	2	Aav52227 Streptoco
7	4163	100.0	110000	7	Continuation (10 o
8	4163	100.0	110000	7	ABSS56454_10

9	3218	77.3	2523	3	AAA5731	3	AAA5731	Streptoco
10	3218	77.3	2647	3	AAA5736	3	AAA5736	Streptoco
11	3218	77.3	2647	6	ABK15103	6	ABK15103	DNA encod
12	2812	67.5	2478	3	AAA08557	3	AAA08557	S. pneumo
13	2797.5	67.2	2457	7	ABX06885	7	ABX06885	S. pneumo
14	2795	67.1	2481	3	AAA05417	3	AAA05417	Streptoco
15	2786.5	66.9	2531	3	AAA47605	3	AAA47605	Recombina
16	2772	66.6	2517	7	ABX06705	7	ABX06705	S. pneumo
17	2772	66.6	2531	3	AAA47602	3	AAA47602	Recombina
18	2772	66.6	110000	7	ABSS56454_08	7	ABSS56454_08	Continuation (9 of
19	2750.5	66.0	2639	3	AAA5737	3	AAA5737	Streptoco
20	2750.5	66.0	2639	6	ABK15104	6	ABK15104	DNA encod
21	2649.5	63.6	2290	6	AAV27356	6	AAV27356	Streptoco
22	2649.5	63.6	2290	6	ABQ84824	6	ABQ84824	S. pneumo
23	2649.5	63.6	2290	9	ADC45146	9	ADC45146	S. pneumo
24	2431.5	58.4	2163	2	AAA25394	2	AAA25394	Streptoco
25	1774	42.6	3159	2	AAV52376	2	AAV52376	Streptoco
26	1591.5	38.2	3171	3	AAA5739	3	AAA5739	Streptoco
27	1247	29.9	3120	3	AAA5730	3	AAA5730	Streptoco
28	1247	29.9	5048	3	AAA5735	3	AAA5735	Streptoco
29	1247	29.9	5048	6	ABK15101	6	ABK15101	DNA encod
30	1246	29.9	3117	7	ABX06706	7	ABX06706	S. pneumo
31	1222.5	29.4	6867	2	AAV52325	2	AAV52325	Streptoco
32	1216.5	29.2	110000	6	ABN71527_18	6	ABN71527_18	Continuation (19 o
33	1203.5	28.9	1455	3	AAA05473	3	AAA05473	Streptoco
34	1203.5	28.9	1455	3	AAZ91804	3	AAZ91804	Streptoco
35	1203.5	28.9	1455	3	AAA5733	3	AAA5733	Streptoco
36	1203.5	28.9	1455	3	AAA47603	3	AAA47603	Recombina
37	1187.5	28.5	1342	2	AAV27414	2	AAV27414	Streptoco
38	1187.5	28.5	1342	6	ABQ84882	6	ABQ84882	S. pneumo
39	1187.5	28.5	1342	9	ADC45262	9	ADC45262	S. pneumo
40	988	23.7	973	2	AAV52488	2	AAV52488	Streptoco
41	966.5	23.2	2528	3	AAA5738	3	AAA5738	Streptoco
42	966.5	23.2	2528	6	ABK15105	6	ABK15105	DNA encod
43	950	22.8	2475	6	ABN66838	6	ABN66838	Streptoco
44	950	22.8	2478	5	AAAS0036	5	AAAS0036	Streptoco
45	946	22.7	5215	3	AAA5740	3	AAA5740	Streptoco
46	946	22.7	5215	3	AAK91105	3	AAK91105	Group B S
47	941	22.6	2472	3	AAA5741	3	AAA5741	Streptoco
48	938	22.5	1398	6	ABN66839	6	ABN66839	Streptoco
49	937	22.5	2466	6	ABN69535	6	ABN69535	Streptoco
50	937	22.5	2466	6	ABN70334	6	ABN70334	Streptoco
51	937	22.5	110000	6	ABN71527_12	6	ABN71527_12	Continuation (13 o
52	934	22.4	2469	5	AAAS0038	5	AAAS0038	Streptoco
53	929	22.3	2469	3	AAA05811	3	AAA05811	Group B S
54	841	20.2	1146	3	AAA05814	3	AAA05814	Group B S
55	770.5	18.5	1684	2	AAV52391	2	AAV52391	Streptoco
56	650	15.6	2721	8	ACF42880	8	ACF42880	S. pneumo
57	644	15.5	816	8	ACF42878	8	ACF42878	S. pneumo
58	644	15.5	816	8	ACF42879	8	ACF42879	S. pneumo
59	643	15.4	819	6	ABK15106	6	ABK15106	DNA encod
60	643	15.4	819	8	ACF42881	8	ACF42881	S. pneumo
61	628	15.1	504	3	AAA08556	3	AAA08556	S. pneumo
62	546	13.1	492	2	AAK25393	2	AAK25393	Streptoco
63	512.5	12.3	1455	3	AAK05838	3	AAK05838	Group B S
64	373.5	9.0	1137	6	ABN66837	6	ABN66837	Streptoco
65	304.5	7.3	841	2	AAAT28529	2	AAAT28529	S. pneumo
66	304.5	7.3	841	4	ABA76858	4	ABA76858	Streptoco
67	302	7.3	2631	6	ABN68134	6	ABN68134	Streptoco
68	302	7.3	110000	6	ABN71527_08	6	ABN71527_08	Continuation (9 of
69	261	6.3	1729	6	AAV42952	6	AAV42952	Streptoco
70	223	5.4	2376	6	ABN68135	6	ABN68135	Streptoco
71	223	5.4	2379	5	AAAS0037	5	AAAS0037	Streptoco
72	221	5.3	567	7	ABX06707	7	ABX06707	S. pneumo
73	176.5	4.2	10242	4	ABL20787	4	ABL20787	Drosophil
74	176.5	4.2	20978	4	ABL20786	4	ABL20786	Drosophil
75	172	4.1	7446	7	ACA19708	7	ACA19708	Prokaryot
76	172	4.1	7446	9	ACF05848	9	ACF05848	Methicill
77	171	4.1	2685	7	ABT14858	7	ABT14858	Pathogen
78	170	4.1	7434	4	AAAS2179	4	AAAS2179	Staphyloc
79	170	4.1	7434	7	ABT14965	7	ABT14965	Pathogen
80	170	4.1	7434	7	ACF73459	7	ACF73459	Staphyloc
81	170	4.1	7437	4	AAAS5232	4	AAAS5232	Staphyloc

82 169 4.1 2685 7 ACF72781 Staphyloc  
83 169 4.1 2730 7 ABZ22901 Staphyloc  
84 169 4.1 4703 7 ADA89766 Staphyloc  
85 168.5 4.0 3132 6 ABK48054 CDNA enco  
86 168.5 4.0 5388 4 ABZ25979 Drosophil  
87 166 4.0 3067 4 AAH54788 S. epider  
88 166 4.0 15071 4 AAH53785 S. epider  
89 164 3.9 2676 7 ACA19807 Prokaryot  
90 163.5 3.9 4505 4 ABL17779 Drosophil

## ALIGNMENTS

## RESULT 1

AAV27351  
ID AAV27351 standard; DNA; 2389 BP.

XX AC AAV27351;

XX DT 02-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae SP0036 nucleotide.

XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
XX detection; pneumonia; otitis media; meningitis; ss.

XX OS Streptococcus pneumoniae.

XX FH Key Location/Qualifiers  
XX CDS 2..2389

XX FT /\*tag= a

XX FT /product= "SP0036"

XX FT /transl\_except= (pos.1367..1369,aa:Xaa)

XX FT /note= "no stop codon given; Xaa is unspecified"

XX PN W09818930-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US019422.

XX PR 31-OCT-1996; 96US-0029960P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;

XX PT WPI; 1998-272224/24.

XX DR P-ESDB; AAW55090.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
XX - or their epitope-containing fragments, useful in protective or  
XX therapeutic vaccines, and for diagnosis.

XX Claim 1; Page 59; 118pp; English.

XX The present sequence encodes a protein from Streptococcus pneumoniae. The  
XX nucleic acid sequence encoding the Streptococcus pneumoniae protein can  
XX be useful in vaccines for inducing protective antibodies against  
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
XX are used to detect Streptococcus infection (by usual hybridisation or  
XX amplification methods), also for isolating Streptococcus genes or their  
XX allelic variants. The protein can be used similarly to detect specific  
XX antibodies in standard immunoassays, especially for diagnosing or  
XX monitoring infections. Antibodies which bind the protein are used to  
XX detect corresponding antigens, to purify the protein and for passive  
XX immunisation (optionally coupled to a toxin). Vaccines are administered,  
XX e.g. by injection, orally or through the skin, typically at 0.01-1000  
XX (especially 10-300) mu g/ml per dose

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

XX SQ

Alignment Scores:  
Pred. No.: 2e-282 Length: 2389  
Score: 4163.00 Matches: 796  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.95% Indels: 0  
DB: 2 Gaps: 0

US-09-765-271-56 (1-796) x AAV27351 (1-2389)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
DB 2 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTCAT 61  
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
DB 62 ATAGATGGAAACCAAGCGACGCAAAAACGAGAAATTTGACTCCTGTAGGTTAGCAAG 121  
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
DB 122 CGTGAGGATCAATGCTGACCAATCGTCATCAAGATAACAGACCAAGGCTATGTCACT 181  
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80  
DB 182 TCACATGCGGACCACTATCATTTACAATGGTAAGGTTCTTATGACGCTATCATCACT 241  
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
DB 242 GAAGAAATTACTCATGAAGATCCAAACTATAAGCTAAAAGATGAGGATTTGTTAATGAG 301  
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
DB 302 GTCAAGGGTGGATATGTTATCAAGGTAGATGAAAATACATATGTTTACCTTAAGGATCT 361  
QY 121 AlaHisAlaAspAsnValArgThrLysGluIleLeuAsnArgGlnLysGlnGluHisSer 140  
DB 362 GCCCAGCGGATAACGTCCTGTACAAAAGAGAAATCAATCGACAAAAACAAGAGCATAGT 421  
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
DB 422 CAACATCGTGAAGGTGGAACCTCCAAGAAACAGTGTGCTGTGTGCTTGGCAGCTTCGAA 481  
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
DB 482 GGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACT 541  
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
DB 542 GGTGATGCTTATATCGTTCTCTCATGGAGATCATTTACCATTTACCTTCTTAAAGATGAGTTA 601  
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
DB 602 TCAGCTAGCGAGTTGGCTGTGTCGAGAAGCCTTCCTATCTGCTCGAGGAAATCTGTCAAT 661  
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
DB 662 TCAAGAACCTATCCCGACAAAATAGCGATAAACAACCTTCAAGAACAACTGGGTACCTTCT 721  
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
DB 722 GTAAGCAATCAGAACTACAAATACAAACAGCAACCAACAGCAACACACTAAACAGTCAA 781  
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
DB 782 GCAAGTCAAGTAAATGACATGATGATAGTCTCTTGAACAGCTCTACAACTGCCCTTGAGT 841  
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
DB 842 CAACGACATGTAGAATCTGATGGCTGTGCTTTGATCCAGCACAAATCACAAAGTCGACA 901  
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
DB 902 GCTAGAGGTGTTGAGTGGCCACAGGAGATCATTTACCACTTCATCCCTTCTCTCAAATG 961

QY 321 SerGluLeuGluGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db TCTGAATTTGGAGAACGAAATCGCTCGTATTAATCCCTTCGTTATCGTTCAACATGG 1021  
 QY 341 ValProAspSerArgProGluGluProSerProGlnProThrProGluProSerProGly 360  
 Db GTACCAGATTCAAGCCAGAACCAAGTCCACAAACCGACTCCGGAACCTAGTCCAGC 1081  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db CGCAACTCGCACCAAAATCTTAAATATGACTCAAAATCTCTCTTTGGTTAGTCAGCTGTA 1141  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db CGAAAAGTTGGGAAGATATGATTCCAGAGAAAGGCACTCTCTGTTAATGCTTTTGGC 1201  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db AAAGATTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 1261  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db GTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTGAACCAAGAAATTTAT 1321  
 QY 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTGTAATAAGGGTCGT 1381  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db AATTCTGATTTCCAAAGCTTAGACAAATATTAGAAGCTTGAATGATGAATCGACTAAT 1441  
 QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db AAAGAAATTTGGTAGATGATTTATTGGCAATCTCTAGCACCAATTAACCCATCCAGACGA 1501  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db CTTGGCAACCAATCTCAAAATGAGTATACTGAGACGAGTTCGTATTGCTCAATTA 1561  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db GCTGATAAGTATACACGTCAGATGTTTACATTTTGTATGAACATGATATAATCAGTAT 1621  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560  
 Db GAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACTGGATTGGAAAGATAGC 1681  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db CTTTCTGATTAAGAAAGTTGACGCTCAAGCCTATATAAGAAAGGTTATCTTACCT 1741  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATC 1801  
 QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db AATCGTGTGAAGGGGAAACCAATTTCCACTCGTTCATATATGTTGGTTAGCAT 1861  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db ACAGTTGAGGTTAAACCGGTAATTTGATTTATCTCTCATGAAGATCATACCATATATT 1921  
 QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db AAAATTTGCTTGGTTTGAATGATACACATACAAAGCTCCAAATGCTATACCTTGGAGAT 1981  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db TTTGTTGCGCAGATTAAAGTACTAGTAGAACACCCCTGACGAACGTCCACATTTCTAATGAT 2041

QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db GGATGGGGCAATGCGCAGTGAGCATGTGTTAGGCAAGAAAGACCACAGTGAAGATCCAAAT 2101  
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAAGAGCAGAAAGTTTTCCTTGGGAAA 2221  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db GTAACGGATTCCTAGTCTGAAAGCAATGCAACACAACTCTAGCTGGTTTTCGAAATTAAT 2281  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db TTGACTCTTCAAAATTATGGATAACAATAGTATATCATGGCAGAGCAGAAATAATTTACTTGGC 2341  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db TTGTTAAAGGAGTAATCCTTCTATCTGTAAGTAAGGAAAAATAAAC 2389  
 RESULT 2  
 ABQ84819  
 ID ABQ84819 standard; DNA; 2389 BP.  
 XX  
 AC ABQ84819;  
 DT 04-SEP-2002 (first entry)  
 XX  
 DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.  
 XX  
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection; gene; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN US2002061545-A1.  
 XX  
 PD 23-MAY-2002.  
 XX  
 PF 22-JAN-2001; 2001US-00765272.  
 XX  
 PR 30-OCT-1997; 97US-00961083.  
 PA (CHOI/) CHOI G H.  
 PA (KUNS/) KUNSCH C A.  
 PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 XX WPI; 2002-479261/51.  
 DR P-PSDB; ABP54584.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
 PT and for preventing or attenuating disease caused by Streptococcus  
 PT infection.  
 XX  
 PS Claim 1; Page 27; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
 CC pneumoniae antigens have antibacterial activity and can be used in  
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
 CC attenuate a Streptococcal infection in an animal. The polynucleotides  
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus

CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning  
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
 CC from the present invention

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 2e-282 Length: 2389  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: 6 Gaps: 0

US-09-765-271-56 (1-796) x AB084819 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 2 TCTTACGAGTTGGGACTGCTCAAGCTAGAACGGTTAAGGAAAATAATCGTGTTCCTAT 61  
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 62 ATAGATGAAACACAGCCAGCCAAACGAGAAATTCACCTCCGATGAGGTAGCAAG 121  
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 122 CQTGAAGAAATCAATGCTGAGCAATCGTCATCAGATTAACAGCAAGGCTATGCTACT 181  
 Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeSer 80  
 Db 182 TCACATGCGCACCATCATATTATTAACATGTAAGTTCCTTATGACGCTATCATCAT 241  
 Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 242 GAAGAAATTAATCATGAAAGATCCAAACATATAAGCTAAAGATGAGATATTTGTAATGAG 301  
 Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120  
 Db 302 GTCAGGGTGGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGCT 361  
 Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnIleHisSer 140  
 Db 362 GCCCACGGGGATACGTCGTCACAAAAGAGGAAATCAATCGACAAACAAAGAGCATAGT 421  
 Qy 141 GlnHisArgGluGlyGlyTyrProArgAsnAspGlyAlaValAlaIleAlaArgSerGln 160  
 Db 422 CAACATCGTGAAGTGGAACTCCAGAAACGATGGTGTGCTTGCCTGGCAGCTTCGCA 481  
 Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 482 GGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGGATCT 541  
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 542 GGTGATGCTTATATGCTTCCATGAGATCAATACCATTAATCTCCTAAGAATGAGTTA 601  
 Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyValGlyAsnLeuSerAsn 220  
 Db 602 TCAGCTAGCGAGTTGGCTGCTCGAAGGCTTCCCTATCTGCTGCGAGAAATCTGTCAAT 561  
 Qy 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 662 TCAAGAACCTATCGCCGACAAATAAGCATAAACATTCACAAACAACTGGGTACCTTCT 721  
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 722 GTAAGCAATCCAGAACTACAATACTAACAAAGCAACAAACAGCAACACTAACACTCAA 781  
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 782 GCAAGTCAAGTAATGACATTGATCTCTTGAACACGCTCTACAAACTGCTCTTGGT 841  
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300

Db 842 CAACGACATGTAGAATCTGATGGCTTGTCTTTGATCCAGCACAAATCACAAAGTCGAACA 901  
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 902 GCTAGAGGTGTGAGTGGCCACACGAGATCATTACCACCTTCATCCCTTACTCTCAATG 961  
 Qy 321 SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 962 TCTGAATTTGGAAGAACGAATCGCTCGTATTATTCCTCTGTTATCGTTCAAACCATGG 1021  
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1022 GTACCAGATTCAAGGCCAGAAACAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1081  
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1082 CCGCAACCTGACCAAAATCTTAAANAGACTCAAAATCTTCTTGGTTAGTCAGCTGTA 1141  
 Qy 381 ArgLysValGlyGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400  
 Db 1142 CGAAAGTTGGGAAGGATATGATTTCGAAGAAAGGCGCATCTCTCGTTATGCTCTTGG 1201  
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 1202 AAGATTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTATATCAAAACAAGAGAGT 1261  
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1262 GTTTCACACACTTTTAACTGCTTAAAGAAATGTTGCTCTCTGACCAAGAAATTTAT 1321  
 Qy 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1322 GATTAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCGT 1381  
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1382 AATTCGTGATTTCCAAAGCCTTAGACAAATTTATTAGAACGCTTGAATGATGAAATCGACTAAT 1441  
 Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1442 AAAGAAAATTTGGTAGATGATTTATTTGGCATCTCTAGCACCAATTTACCCATCCAGAGCA 1501  
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1502 CTTTGCAAAACCAATTTCTCAATTCAGTATATCTGAAGACGAAGTTCGTATTGCTCAATTA 1561  
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1562 GCTGATAAGTATACAACTGATGTTTACATTTTGTGTAACATGATATAATCATGATGAT 1621  
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 1622 GAAGAGATGATATGTAACGCTCATATGGCCATAGTCACTGGATGGAAAGATAGC 1681  
 Qy 561 LeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1682 CTTTCTGATAAGCAAAAGTTGCGCTCAAGCCTATCTATAAGAAAAAGGATCTCTACCT 1741  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1742 CCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCACAGACTATTAC 1801  
 Qy 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1802 AATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTCGATCTCCATATATGTTGAGCAT 1861  
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1862 ACAGTTGAGTTAAAAACGGTAAATTTGATTTATTCCTCATAGGATCATACCAATAATTT 1921  
 Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660

Db 1922 AAATTTGCTTGGTTGATGATCATACATCAAAAGCTCCAAATGGCTATACCTTGGAGAT 1981  
 QY 661 LeuPheAlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 1982 TTGTTTGGACGATTAACTACTAGTAGAACACCCCTGAGCAAGCTCCACATTTCTAATGAT 2041  
 QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2042 GGATGGGGCAATGCCAGTGGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2101  
 QY 701 LysAsnPhelLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2102 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTCTGAGCCAGAGTCCCT 2161  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2162 CAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGACAGAAAGTTTGTCTTGGAAA 2221  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2222 GTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT 2281  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 2282 TTGACTCTTCAAAATTATGGATTAACAATAGTATCATGTCAGACAGAGCAAAAATTTACTTGG 2341  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 2342 TTGTTAAAGGAGTAATCTTCTATCTGTAAGTAAGGAAAAATAAC 2389

RESULT 3  
 ID ADC45136  
 AC ADC45136;  
 DT 18-DEC-2003 (first entry)  
 DE S. pneumoniae DNA encoding antigen SP036.  
 KW Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.  
 OS Streptococcus pneumoniae.  
 PF US6573082-B1.  
 PN 03-JUN-2003.  
 PD 28-MAR-2000; 2000US-00536784.  
 PP 31-OCT-1996; 96US-0029960P.  
 PR 30-OCT-1997; 97US-00961083.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX WPI; 2003-764574/72.  
 DR P-FSDB; ADC45137.  
 PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides  
 PT useful for producing vaccines for prevention or attenuation of infection  
 PT by Streptococcus pneumoniae.  
 XX Example 1; SEQ ID NO 55; 58pp; English.  
 CC The invention relates to an isolated polynucleotide consisting of a  
 CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding  
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae  
 CC antigens. Also included are making a recombinant vector by inserting the  
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at  
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a

CC recombinant host cell comprising the SP028 polynucleotide. The nucleic  
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae  
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae  
 CC antigen nucleic acids are useful as probes for use in diagnostic methods  
 CC for detecting S. pneumoniae gene expression. The present sequence encodes  
 CC an S. pneumoniae antigenic protein.

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 0 U; 1 Other;

Alignment Scores:  
 Pred. No.: 2e-282 Length: 2389  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: Gaps: 0

US-09-765-271-56 (1-796) x ADC45136 (1-2389)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 2 TCTTACGAGTTGGACTGTATCAAGCTAGAACCGTTAAGGAAATATATCGTGTTCCTAT 61  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 62 ATAGATGGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGTTAGCAAG 121  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 122 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCACT 181  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 182 TCACATGGCGACCACTATCATATTATCAATGGTAAGGTTCTTATGACGCTATCATCAGT 241  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 242 GAAGATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG 301  
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 302 GTCGAAGGCTGATATGTTATCAAGTAGATGGAATAACTATGTTTACCTTAAGGATGCT 361  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db 362 GCCCAGCGGATTAACGTCGCTACAAAAGAGAAATCAATCGACAAAACAGAGCATAGT 421  
 QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 422 CAACATCGTGAAGTGGAACTCCAAAGAAACGATGCTGCTTGGCAGCTTCGCAA 481  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 482 GGACGCTATACACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGGATACT 541  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 542 GGTGATGCTTATATCGTTCTCATGGAGATCATTTACCATTACATTCCTTAAGAAAGAGTTA 601  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 602 TCAGCTAGAGATTGGCTGCTGAGAGCCCTTCTATCTGTCGAGGAAATCTGTCAAAAT 661  
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 Db 662 TCAAGAACCTATCGCCGACAAAATAGCGATAACACTTCAAGAACAAACTGGGTACTTCT 721  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 722 GTAAGCAATCCAGGAACCTACAAAATACAAAGCAAGCAACAGCAACACTTAACGTCAA 781  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280

Db 782 GCAAGTCAAGTAATGACATTTGATAGTCTCTTGAAACAGCTCTACAACTGCTTTGAGT 841  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 842 CAACGACATGTAGAATCTGATGCTTGTCTTTGATCCAGCACAAATACAAAGTCGAACA 901  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 902 GCTAGAGGTTGTGACGTGCACACGAGATCATTACCACTTCATCCCTTACTCTCAATG 961  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 962 TCTGAATTCGAAGACGAATCGTCTGATATTATTCCTTCTGTTATCCCAACCACTTGG 1021  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1022 GTACCAAGATTCAAGGCCAGAACCAAGTCCACACCGACTCCGGAACCTAGTCAGGC 1081  
 QY 361 ProGlnProAlaProAsnLeuIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1082 CCGCAACCTGCACCAATCTTAAATPAGACTCAAAATCTCTTCTGTTAGTCAGCTGTA 1141  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1142 CGAAGAATTCGGAGAGATATGATTTCGAAGAAAGGGCATCTCTCGTTATGCTTTGGC 1201  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 1202 AAAGATTTACCATCTGMAACTGTAAATCTTGAAGCAAGTTATCAAAACAAGAGAGT 1261  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1262 GTTTCACACACTTTAACTGCTAAAAAGAAAAATGTTGCTCTCTGTCACCAAGAAATTTAT 1321  
 QY 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1322 GATAAAGCATATATCTGTTAACTGAGGCTCAAAAGCCTTGTGTTGNAATTAAGGTCGT 1381  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1382 AATCTGATTTCCAAAGCTTAGACAAATTAATTAGAAGCTTGATGAATGATCGACTAAT 1441  
 QY 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1442 AAAGAAAAATTTGGTAGATGATTTATTTGTCATCTTAGCACCATAATACCCATCCAGAGGA 1501  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1502 CTGGCAACCAATCTCAATTCAGTATATCTGAAGACGAAGTTCTGTTATGCTCAATTA 1561  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1562 GCTGATAAGTATACAACTCAGATGTTTACATTTTGTATGAACATGATATATATCAGTAT 1621  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 1622 GAAGAGATGCATATGTAACCCCTCATATGGGCCATAGTCACTGGATTGGAAAAAGATAGC 1681  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1682 CTTTCTGATAAGAAAGATTGAGCTCAAGCCTATATAAGAAAAAGGATCTCTACCT 1741  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1742 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1801  
 QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1802 AATCTGTGAAGGGGAAAAACGAATTCATCTGTTGACTTCCATATATATGTTGAGCAT 1861  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1862 ACAGTTGAGTTAAAAACGGTAATTTGATTATCTCTATAAGGATCATTAACATAATATT 1921

QY 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1922 AAATTTGCTTGGTTTGTATGATCACATACAAAGCTCCAAATGGCTATACCTTTGGAAGAT 1981  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 1982 TTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGACGTCCACATCTTAATGAT 2041  
 QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2042 GGATGGGGCAATGCCAGTGCAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAAT 2101  
 QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2102 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAACACCTGTGAGCCAGAGTCCCT 2161  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAGTTTGTCTGGGAA 2221  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2222 GTAACGGATTTAGTCTCAAGCCCAATGCACAGAACTCTAGCTGGTTTACGAAATAAT 2281  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 2282 TTGACTCTTCAATTTATGGATTAACATAGTATCATGGCAGAGCAGAAAAATTACTTGGC 2341  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 2342 TTGTTAAAGGAAGTAATCTTCACTCTGTAGTAAGGAAAAATAAAC 2389  
 RESULT 4  
 ABX06886  
 ID ABX06886 standard; DNA; 2406 BP.  
 AC ABX06886;  
 DT 27-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain coding region #1174.  
 XX  
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR P-PSDB; ABU01598.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 6; SEQ ID NO 2347; 56pp; English.



CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC AB56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2489  
 CC identified coding region from the genomic sequence. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 CC standardise OS field)

XX Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2,01e-282 Length: 2406  
 Score: 4163.00 Matches: 795  
 Percent Similarity: 99.87% Conservative: 0  
 Best Local Similarity: 99.87% Mismatches: 1  
 Query Match: 99.95% Indels: 0  
 DB: 7 Gaps: 0

US-09-765-271-56 (1-796) x ABX06886 (1-2406)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 DB 19 TCTTACGAGTGGGACTGTATCAAGCTAGACGGTTAAGGAAATAATCGTGTTCCTAT 78  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 DB 79 ATAGATGGNAACACGACGCGAATAACGGAGAAATTCACCTCCTGATGAGGTAGCAAG 138  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValLysLysLysLysLysLysLysLysLys 60  
 DB 139 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGCAAGGCTATGTCACT 198  
 QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleLysSer 80  
 DB 199 TCACATGGCGACCATATCATATTACAAATGTAAGGTTCCTATGACGCTATCATCAT 258  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 DB 259 GAAGAATTTACTCATGAAAGATCCAACTATATAGCTAAAGATGAGATATTGTTATGAG 318  
 QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 DB 319 GTCAAGGGTGGATATCTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGGATGCT 378  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140  
 DB 379 GCCACGCGGATACCTCGGTACAAAGAGGAATCAATCGACAAACAAAGAGCATAGT 438  
 QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyValAlaLeuAlaArgSerGln 160

DB 439 CAACATCGTGAAGGTGGAACTCCAAAGAACGATGGTGTGCTTGGCTTGGCGCTTGGCAA 498  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 DB 499 GGACGCTATACATACAGATGATGGTTATATCTTATATGCTTCTGATATCATAGAGATACT 558  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 DB 559 GGTGATGCTTATATGCTTCCATCGGAGATCATTAACCATTTACATTCCTTAAGATGAGTTA 618  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 DB 619 TCAGCTAGCGAGTTGCTGCTGCAGAGCCTTCTCTATCTGGTCGAGGAAATCTGTCAAAT 678  
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 DB 679 TCAAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACCAAACTGGGTACTCTCT 738  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 DB 739 GTAAGCAATCCAGGAACCTACAAATACTAACCAAGCAACACAGCAACACTAACAGTCAA 798  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 DB 799 GCAAGTCAAAAGTAATGACATGATAGTCTCTTTGAAACAGCTCTACAAACCTGCTTTGAGT 858  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 DB 859 CAACGACATGTAGAATCTGTCTTTGATCCAGCACAAAATCAACAGTCCGAAACA 918  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 DB 919 GCTAGAGGTGTTCAGTGCCACACGAGAGATCATTAACCACTTCATCCCTTACTCTCAATG 978  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 DB 979 TCTGAATTTGGAAGAACGAATCGTCTGTTATTTCCCTTCGTTATCGTTCAAAACCATGG 1038  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 DB 1039 GTACCAGATTTCAAGGCCAGAACCAACCAAGTCCACCAACCGACTCCGGAACCTAGTCCAGGC 1098  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 DB 1099 CCGCAACCTCGCACCAAACTTTAAATAGACTCAATTTCTTTTGGTGTAGTCAGTGGTA 1158  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 DB 1159 CGAAAAGTTGGGGAAGGATATGTTATTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGGC 1218  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysSerLysGlnGluSer 420  
 DB 1219 AAAAGATTTTACCCTATGAAACCTGTTAAATACTTTGAAAGCAAGTTATTCAAAACAAAGAGT 1278  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 DB 1279 GTTTTACACACTTTTAACTGCTAAAAAGAAAATGTTGCTCTCTCGTACCAAGATTTTAT 1338  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 DB 1339 GATAAAGCATATAATCTGTGTTAACTGAGGCTCATAAAGCCCTGTTTGAATAAAGGGTCGT 1398  
 QY 461 AsnSerAspPheGluAlaLeuAspLysLeuGluArgLeuAsnAspGluSerThrAsn 480  
 DB 1399 AATTCCTGATTTTCCAGCCCTTAGACAAATTTATTAGAACCGCTTGAATGATGAATCGCATAT 1458  
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 DB 1459 AAAAGAAAATTCGTAGATGATTTTATTTGGCATTTCTTAGACCAATTAACCATCCAGAGCGA 1518  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 DB 1519 CTTGGCAACCAAAATTTCTCAATTTGAGTATATCTGAAGACGAAGTTCTGTTATGCTCAATTA 1578

XX	<i>Streptococcus pneumoniae</i> : infection; vaccine; coiled coil region;
XX	<i>Streptococcus pneumoniae</i> : infection; vaccine; coiled coil region;
KW	histidine triad residue; Sp36; antibody; otitis media;
KW	nasopharyngeal infection; bronchial infection; bronchitis; sepsis
KW	meningitis; lobar pneumonia; ds.
XX	
XX	<i>Streptococcus pneumoniae</i> .
OS	
OS	<i>Streptococcus pneumoniae</i> .



23-OCT-1998 (first entry)  
 Streptococcus pneumoniae genome fragment SEQ ID NO:94.  
 Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 computer readable medium; vaccine; pharmaceutical composition; ds.  
 Streptococcus pneumoniae.  
 WO9818931-A2.  
 07-MAY-1998.  
 30-OCT-1997; 97WO-US019588.  
 31-OCT-1996; 96US-0029960P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Pannon M;  
 Dougherty BA;  
 WPI; 1998-272225/24.  
 Computer-readable medium with recorded Streptococcus pneumoniae  
 polynucleotide sequences - useful in diagnostic kits and assays, and  
 pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
 Claim 1; Page 727-732; 1409pp; English.  
 The present invention describes a computer readable medium which has the  
 nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 on it, or a representative fragment or a sequence at least 95% identical  
 to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 pneumoniae. The present invention also describes an isolated nucleic acid  
 molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 by a process comprising: (a) screening a genomic DNA library using as a  
 probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 391, identifying members of the library which contain sequences that  
 hybridize to the target sequence and isolating the nucleic acid molecules  
 from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 homologous to amplification primers derived from the fragment of the S.  
 pneumoniae genome to prime the amplification and isolating the amplified  
 sequences. The computer readable medium can be used in a computer-based  
 system for identifying fragments of the S. pneumoniae genome of  
 commercial importance, or expression modulating fragments of the S.  
 pneumoniae genome. Products from the present invention can be used in  
 diagnosis kits and assays, and pharmaceutical compositions and vaccines  
 for S. pneumoniae  
 SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;  
 Alignment Scores:  
 Pred. No.: 8,76e-282 Length: 8195  
 Score: 4163.00 Matches: 795  
 Percent Similarity: 99.87% Conservative: 0  
 Best Local Similarity: 99.87% Mismatches: 1  
 Query Match: 99.95% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-765-271-56 (1-796) x AAV52227 (1-8195)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 3054 TCTTACGAGTTGGAGCTATCAAGCTAGAACGGTTAAAGGAAAATAATCGTGTTCCTAT 3113  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 3114 ATAGATGGGAAAACGACGCAAAACCGAGAAATTTGACTCCTGATGAGGTTAGCAAG 3173

41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 3174 CGTGAAGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGGCTATGCTCACT 3233  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80  
 Db 3234 TCACATGCGCACCACTATCATATTACAAATGTAAGGTTCCTTATGACGCTATCATCACT 3293  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 3294 GAAGAAATTACTCATGAAAGATCCAAATATTAAGCTTAAAGATGAGATATGTTAATGAG 3353  
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValLysAspAla 120  
 Db 3354 GTCAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACTTAAAGATGCT 3413  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140  
 Db 3414 GCCCAGCGGATACCGTCAAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 3473  
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 3474 CAACATCGTGAAGGTGGAACTCCAAAGAAACGATGGTGTGCTTGGCCTTGGCAGTTCCGAA 3533  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 3534 GGAAGCTATACTACAGATGATGTTTATATCTTAAATGCTTCTGATATCATAGAGTACT 3593  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 3594 GGTGATGCTTATATCGTCTCATGGAGATCAATACCATTCATCTCTAAGAATGAGTTA 3653  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 3654 TCAGCTAGGAGTTGGCTGCTGCAGAGCCCTTCTATCTGGTCGAGGAAATCTGTCAAAT 3713  
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 3714 TCAAGAACCTATCGCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTTCT 3773  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 3774 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 3833  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 3834 GCAAGTCAAAAGTAATGACATTCATGCTCTTTGAAACAGCTCTACAAACTGCTTTGAGT 3893  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 3894 CAACGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGCACAAATCAAGTCGAACA 3953  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 3954 GCTAGAGGTGTGACGTGCCACGAGATCAATTCACCTTCTATCCCTTACTCTCAATG 4013  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 4014 TCTGAATTCGAAGACGAATCGCTGCTATTTATTCCTCTTTCCTTATCGTTCAAAACCATGG 4073  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 4074 GTACCAAGATTCAGGCGCAGAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4133  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 4134 CCGCAACCTGCGACCAATCTTAAATAGACTCAAATCTCTTGTGTAGTCAGCTGGTA 4193  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 4194 CGAAAAGTTGGGAGGATATGTTATCGAAGAAAAGGAGCATCTCTCTGTTATGCTTTGCG 4253  
 QY 401 LysAspLeuProSerSerGluThrValLysAsnLeuGluSerLysGlnGluSer 420

```
Db 4254 AAAAGATTACCATCTGAACATGTTAAAAATCTTGAAACCAAGTATTCAAAACAGAGAGT 4313
Qy 421 ValSerHisThrLeuThrAlaLysGluValAlaProArgAspGlnGluPheTyr 440
Db 4314 GTTTTACACACATTTAACTGCTTAAAAAGAAAAATGTTGCTCTCGTACCAAGAATTTTAT 4373
Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 4374 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCCTGTTTGAATAATTAAGGGTCGT 4433
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 4434 AATCTGATTTCCAAAGCTTAGACAAATATTAGAACCTTGAATGATGAATCACTAAT 4493
Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProLysHisProGluArg 500
Db 4494 AAAGAAAAATGGTAGATGATTTATTGGCATTTCTAGCACCAATTAACCCATCCAGAGCGA 4553
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 4554 CTTGGCAAAACCAAAATCTCAAAATGAGTATACCTGAAGACGAAGTTCGTATTGCTCAATTA 4613
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540
Db 4614 GCTGATAAGTATACAACTGATGATGTTTACATTTTGTGATGACATGATATATCACTGAT 4673
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisThrIleGlyLysAspSer 560
Db 4674 GAAGGAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGGAATGGAAAAAGATAGC 4733
Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580
Db 4734 CTTTCTGATGAAGAAAAAGTTGACGCTCAAGCCTATATCAAGAAAAAGGTATCTCTACCT 4793
Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600
Db 4794 CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGAGCAGCTATTATAC 4853
Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
Db 4854 AATCGTGTGAAGGGGAAAAACGAATCCCACTCGTTCATATATATGTTGAGCAT 4913
Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
Db 4914 ACAGTTGAGTTTAAAAACGGTAATTTGATTTATCTCTCATAGGATCATTAACCAATAATATT 4973
Qy 641 LysPheAlaThrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
Db 4974 AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATATCTTGGAAAGAT 5033
Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
Db 5034 TTGTTTTCGACGATTAAGTACTAGTACGTAAGAACACCCCTGACGAACGCTCCACATTTCTAATGAT 5093
Qy 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
Db 5094 GGATGGGCAATGCCAGTGAGCATGTTTAGGCAAGAAAGAACACACAGTGAAGATCCAAAT 5153
Qy 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720
Db 5154 AAGAACTTCAAGCGGATGAAGCCAGTAGAGGAACACCTGCTGAGCCAGAGTCCCT 5213
Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db 5214 CAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGAACGAGAGTTTGTTCGCGAAA 5273
Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db 5274 GTAAAGGATTTCTAGTCTGAAGGCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 5333
Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
```

```
Db 5334 TTGACTCTTCAATATATGGATAACAATAGTATCATGGCAGAACGAGAAAAAATTACTTCGC 5393
Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 5394 TTGTTAAAAAGGAAGTAATCTTTCATCTGTAAGTAAGGAAAAAATAAAC 5441

RESULT 7
ABS56454_09/c
Continuation (10 of 22) of ABS56454 from base 900001 (Streptococcus pneumoniae type 4 str)
WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454
WP Fragment Name Begin End
WP ABS56454_00 1 110000
WP ABS56454_01 100001 210000
WP ABS56454_02 200001 310000
WP ABS56454_03 300001 410000
WP ABS56454_04 400001 510000
WP ABS56454_05 500001 610000
WP ABS56454_06 600001 710000
WP ABS56454_07 700001 810000
WP ABS56454_08 800001 910000
WP ABS56454_09 900001 1010000
WP ABS56454_10 1000001 1110000
WP ABS56454_11 1100001 1210000
WP ABS56454_12 1200001 1310000
WP ABS56454_13 1300001 1410000
WP ABS56454_14 1400001 1510000
WP ABS56454_15 1500001 1610000
WP ABS56454_16 1600001 1710000
WP ABS56454_17 1700001 1810000
WP ABS56454_18 1800001 1910000
WP ABS56454_19 1900001 2010000
WP ABS56454_20 2000001 2110000
WP ABS56454_21 2100001 2162598

Alignment Scores:
Pred. No.: 1 98e-280 Length: 110000
Score: 4163.00 Matches: 795
Percent Similarity: 99.87% Conservative: 0
Best Local Similarity: 99.87% Mismatches: 1
Query Match: 99.95% Indels: 0
DB: 7 Gaps: 0

US-09-765-271-56 (1-796) x ABS56454_09 (1-110000)
Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 106949 TCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAAGGAAATATCTGTTTCTAT 106890
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 106889 ATAGATGGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAAG 106830
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 106829 CGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATAACAGACCAAGCTATGCTACT 106770
Qy 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 106769 TCACATGGCGACCACTATCATTTATTAATGTTAAGTTCTTCTTATGACGCTATCATCAGT 106710
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 106709 GAAGAATTAATCTCATGAAGATCCAAACTATAAGTAAAAAGATGAGGATATTGTTAATAG 106650
Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 106649 GTCAGGGTGGATATGTTTATCAAGGTAGATGGAATACTATGTTTACCTTAAAGGATGCT 106590
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 106589 GCCCAGCGGATAACGTCGCTACAAAGAGGAATCAATCGACAAAAACAAGAGCATAGT 106530
Qy 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
```

Db	105449	CTTGGCAAAACCAAAATTTCTCAAATTTGAGTATATCTGAAGACGAAGTTCTGTTGCTCAATTA	105399
Qy	521	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp	540
Db	105389	GCTGATTAAGTATACACAGTCAGATGGTTACATTTTGTATGAACATGATATATCAGTGAT	105330
Qy	541	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer	560
Db	105329	GAAGGAGATGCATATGTAAACGCTCATATGGCCCATAGTCACTGGATTTGGAAAAAGATAGC	105270
Qy	561	LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLysLeuPro	580
Db	105269	CTTTCTGATAAGGAAAAAGTTGCGACTCAAGCCTATATCAAGAAAAAGTATCTTACT	105210
Qy	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr	600
Db	105209	CCATCTCCAGACGCGAGATGTTAAAGCAATCCAACCTGGAGATAGTGCAGCAGCTATTAC	105150
Qy	601	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
Db	105149	AATCGTGTGAAGGGGAAAAACGAATTCCTGCTCGTTCCATATATATGTTGAGCAT	105090
Qy	621	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
Db	105089	ACAGTTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATPAAGGATCATTTACCATTAAT	105030
Qy	641	LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
Db	105029	AAATTTGCTTGGTTTGTATGATCACAATACACATACAAAGCTCCAAATGGCTATACCTTGAAGAT	104970
Qy	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	104969	TTGTTTGGCAGCATTAAGTACTACGTAGAACACCCCTGACGAAACGTCACATTTCTAATGAT	104910
Qy	681	GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
Db	104909	GGATGGGCAATGCCAGTGAGCATGTGTAGCCAAAGAACCCACAGTGAAGATCCAAAT	104850
Qy	701	LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro	720
Db	104849	AAGAACTTCAAGCCGATGAAGCCAGCTAGAGGAAACACCTGTCTGAGCCAGAGTCCCT	104790
Qy	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
Db	104789	CAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAAGCAGAAAGTTTGTCTGCGAAA	104730
Qy	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
Db	104729	GTAACGATTTCTAGTCTGAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT	104670
Qy	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
Db	104669	TTGACTTTCAAATTTATGGATAACAATAGTATCATGGCAGAGACAGAAAAATTTACTTGGC	104610
Qy	781	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
Db	104609	TTGTTTAAAGGAAGTAATCTTTCATCTGTAAAGTAAAGAAAAATAAAC	104562
RESULT 8			
ABSS56454_10/c			
Continuation (11 of 22) of ABSS56454 from base 1000001 (Streptococcus pneumoniae t			
WP	Sequence split into 22 fragments	LOCUS	ABSS56454 Accession
WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000



WP	ABS56454_10	1000001	1110000	
WP	ABS56454_11	1100001	1210000	
WP	ABS56454_12	1200001	1310000	
WP	ABS56454_13	1300001	1410000	
WP	ABS56454_14	1400001	1510000	
WP	ABS56454_15	1500001	1610000	
WP	ABS56454_16	1600001	1710000	
WP	ABS56454_17	1700001	1810000	
WP	ABS56454_18	1800001	1910000	
WP	ABS56454_19	1900001	2010000	
WP	ABS56454_20	2000001	2110000	
WP	ABS56454_21	2100001	2162598	

Alignment Scores:

Pred. No.: 1,98e-280 Length: 110000

Score: 4163.00 Matches: 795

Percent Similarity: 99.87% Conservative: 0

Best Local Similarity: 99.87% Mismatches: 1

Query Match: 99.95% Indels: 0

DB: 7 Gaps: 0

US-09-765-271-56 (1-796) x ABS56454\_10 (1-110000)

QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAsnAsnArgValSerTyr	20
Db	6949	TCITACCAGTGGGACGTGTATCAACGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTAT	6890
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	6889	ATAGATCGAAAAAANAAGCGACGCAAAAAAAGGAGAAATTGTACTCTGTAGGTAGTTAGCAAG	6830
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Db	6829	CGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGCGTATGTCACT	6770
QY	61	SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
Db	6769	TCATATGGCACCACATCATCATATTATCAATGGTAAAGTTCCTTATGACGCTATCATCAGT	6710
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	6709	GAAGATTACTCATGAAGATCCAACTAATAGCTAAAGATGAGGAAATTTGTTATGAG	6650
QY	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	6649	GTCAAGGTCGGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGGATGCT	6590
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140
Db	6589	GCCACGCGGATACGTCCTGTACAAAAGAGAAATCAATCGCAAAAACCAAGAGCATAGT	6530
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160
Db	6529	CAACATCGTGAAGGTGGAATCCCAAGAAACGATGGTGCTGTGCTTGGCACGTCGCCAA	6470
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	6469	GGACGCTATACTACAGATGATGGTTATATCTTTAAATGCTTCATATCATAGAGGATACT	6410
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	6409	GGTGATGCTTATATPGTTCCTCATGGAGATCATTAACATTACATTCTCTAAGAATGAGTTA	6350
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	6349	TCAGCTAGCGAGTTGGCTGCTGCAGAGCCCTCTATCTGGTCGAGAAATCTGTCAAAAT	6290
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
Db	6289	TCAAGAACCCTATCGCCGACAAAATAGCGATAACACTTCAAGAACCAACTGGGTACCTTCT	6230
QY	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260



QY 621 ThrValGluValLysAsnGlyAsnLeuLeuLeuProHisLysAspHisTyrHisAsnIle 640  
 Db 5089 ACAGTTGAGGTAAACACGGTAATTTGATTAATCTCTATAAGGATCATTAACCAATATT 5030  
 QY 641 LysPheAlaTTPheAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 5029 AAATTTGCTTGTTGATGATCACATACAAAGCTCCAAATGGCTATACCTTGGAGAT 4970  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 4969 TTGTTTGGCAGGATTAAGTACTACGTAGACACACCTGACGAACGTCACATCTTAATGAT 4910  
 QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 4909 GGATGGGCAATGGCAGTGAGCATGTGTAGGCAAGAAAGACCAAGTGAAGATCCCAAT 4850  
 QY 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 4849 AAGAACTTCAAGCGGATGAGAGCCAGTACAGGAAACACCTGCTGAGCCAGAGTCCCT 4790  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlalys 740  
 Db 4789 CAAGTAGAGACTGAAAGAGTAGAGCCCAACTCAAGAGAGCAGAGTGTTCCTTGCAGAA 4730  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 4729 GTAACGGATTCAGTCTGAAGCCCAATGCACAGAACTCTAGCTGGTTTACGAATTAAT 4670  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 4669 TTGACTCTTCAATTAATGATAAACAATAGTATCATGGCAGAGCAGAAAAATTAATTGGC 4610  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 4609 TTGTTAAAGGAAGTAATCTTCACTGTAGTAAGGAAAAATAAAC 4562

## RESULT 9

AAA65731

ID AAA65731 standard; DNA; 2523 BP.

AC AAA65731;

XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.

DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal; ds.

XX Streptococcus pneumoniae.

OS WO200039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA001219.

XX 23-DEC-1998; 98US-0113800P.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.

XX P-PSDB; AAB12716.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otitis media, bacteraemia and/or pneumonia.

XX Example 2; Fig 3; 106pp; English.

XX

CC The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have  
 CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence encodes the *S. pneumoniae* BVH-11 protein  
 CC antigen

XX  
 SQ Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 3.81e-216 Length: 2523  
 Score: 3218.00 Matches: 616  
 Percent Similarity: 82.95% Conservative: 65  
 Best Local Similarity: 75.03% Mismatches: 102  
 Query Match: 77.26% Indels: 38  
 DB: 3 Gaps: 3

US-09-765-271-56 (1-796) x AAA65731 (1-2523)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 61 GCTTATGAACTAGGTTTGCATCAAGCTCAAACTGTAAAGAAATAATCGGTTCCTAT 120  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 121 ATAGATGAAACAAGACGACGCAAAACGAGGAAATTGACTCTCTGATGAGGTAGCAAG 180  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 181 CGTAGGAATCAACGCCGAACAAATCGTCATCAAGATTACGATCAAGGTTATGTGACC 240  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 241 TCTCATGGAGACCATTAATCTATTAATGCAAGGTCCTTATGATGCCATCATCACT 300  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 Db 301 GAAGAGCTCTCATGAAGAGTCGGAATATCAGTTGAAGATTACAGATTCATGTAAGAA 360  
 QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValTyrLeuLysAspAla 120  
 Db 361 ATCAAGGGTGTTATGTCTAATTAAGGTAAACGGTAAATATCTATGTTTACCTTAAGGATGCA 420  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db 421 GCTCATGGGATAATGTCGTCACAAAAGAAATAATCAATCGCAAAAACAGAACATAGT 480  
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 481 CAGCATCTGAGGAGGAGGACTTCAGCAAAACGATGTCGGTAGCCTTTGACGTTACACAG 540  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 541 GGACGCTACACCAAGATGATGGTTATATCTTCAATGTCATCTGATATCATCAAGATACG 600  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 601 GCGGATGCTATATCTTCCTCATGGAGATCATTAACCATTAATCTCTCAAGATGAGTTA 660  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 661 TCAGCTAGCGAGTTGGCTGTCAGAGCCCTTCCTATCTGTCGGGAAAATCTGTCAAAAT 720  
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 721 TTAAGAACCTATCGCCGACAAAATAAGCGATAACACTCCAAAGAACAAACTGGGTACTCTCT 780  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 781 GTAAGCAATCCAGGAACCTACAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 840

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 DB 841 GCAAGTCAAGTAATGACATGATAGTCTCTTGAAACAGCTCTCAAACTGCGCTTTGAGT 900  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 DB 901 CAACGCCATGTAGATCTGATGGCTTATTTTCGACCCAGCGCAATCACAGTTCGAAC 960  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 DB 961 GCCAGAGTGTAGTGTCTCCCTCATGTTGATGATGATGATGATGATGATGATGATG 1020  
 QY 321 SerGluLeuGluGluArgGlyAlaArgIleLeuProLeuArgTyrArgSerAsnHisTyr 340  
 DB 1021 TCTGAATGGAAACGAATGCTCGTATATTCCTTCCTTCCTTCCTTCCTTCCTTCCT 1080  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 DB 1081 GTACCAAGTTCAGACCAAGAGAACCAAGTCCACACCGACTCCAGAACCTTAGTCCAAGT 1140  
 QY 361 ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374  
 DB 1141 CCGCAACTGCACCAATCTCAACAGCTCCAGCAATCAATGAT-----GAGAAA 1194  
 QY 375 LeuValSerGlnLeuValArgLysValGlyTyrValPheGluGluLysGlyIle 394  
 DB 1195 TTGTCACAAAGAGCTGTTCCGAAAGTAGGCGATGTTATGTCCTTTGAGGAGAAATGGAGTT 1254  
 QY 395 SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuSerLys 414  
 DB 1255 TCTCGTTATATCCAGAGCAAGAAATCTTTACGACGAAACAGCAGCAGCAATGATAGCAA 1314  
 QY 415 LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434  
 DB 1315 CTGCCAAGCAGGAAAGTTATCTCAAGTAGGCTAGGCTAAGAAACCTGACCTCCCATCT 1374  
 QY 435 ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454  
 DB 1375 AGTGATCGAGATTTTACATAGGCTATGACTTACTAGCAAGAAATCCACAGATTTA 1434  
 QY 455 Phe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu 474  
 DB 1435 CTTGATAATAAAGTCGACAAAGTGTATTTGAGGCTTTGGATACCTGTTGGAGCAGCTC 1494  
 QY 475 AsnAspGluSerThrAsnLysGluLeuValAspAspLeuLeuAlaPheLeuAlaPro 494  
 DB 1495 AAGGATGCTCAAGTGATAAGTCAAGTTAGTGGATGATATTTGCTCTTTAGCTCCG 1554  
 QY 495 IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu 514  
 DB 1555 ATTCGTCATCCAGAACGTTTAGGAAACCAATCGCAATATTACCTACATGATGATGAG 1614  
 QY 515 ValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGlu 534  
 DB 1615 ATTCAGTAGTCCCAAGTTGGCAGGCAAGTACACACAGACGCTTATATCTTTGATCCT 1674  
 QY 535 HisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis 554  
 DB 1675 CGTGATATAACCATGATGAGGGGATGCTTGTAACTCCATATGACCCCATAGCCAC 1734  
 QY 555 TrpIleGlyLysAspSerLeuSerAspLysValAlaAlaGlnAlaTyrThrLys 574  
 DB 1735 TGGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCTAA 1794  
 QY 575 GluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAsp 594  
 DB 1795 GAGAAAGGTTTGACCCCTCTCCAGACAGACCATCAGGATTCAGAAATCTGAGGCAAAA 1854  
 QY 595 SerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeu 614  
 DB 1855 GGAAGCAGAGCTATCTACAAACCGCGTGAAGAGCAGCTAAGAAAGGTGCCACTTGATG 1914

QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleProHisLys 634  
 DB 1915 CTTTACCAATCTTCAATATATCTGTAGAGTCAAAAACCGTAGTTTAATCATACCTCATAT 1974  
 QY 635 AspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsn 654  
 DB 1975 GACCATTACCAATACATCAAAATTTGAGTGGTTTTCAGGAAGCGCTTTATGAGCGCACTAAG 2034  
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu 674  
 DB 2035 GGTATATCTCTTGAGGATCTTTTGGCGACTCTCAAGTACTATGTCGAACATCCCAACGAA 2094  
 QY 675 ArgProHisSerAsnAspGlyTyrPheAsnAlaSerGluHisValLeuGlyLysLysAsp 694  
 DB 2095 CGTCCGCAATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAAGAAACAAAAAT 2154  
 QY 695 HisSerGluAspProAsnLysAsnPheLysAlaAsp----- 706  
 DB 2155 GGTCAAGCTGATACCAATCAAAACGAAACCAAGCGGAGAAACCTCAGACAGAAAAA 2214  
 QY 706 ----- 706  
 DB 2215 CCTCAGGAAGAAACCCCTCGAGAAAGAAACCAAAAGCGAGAACACAGAGTCTCCAAA 2274  
 QY 707 -----GluGluProValGluThrProAlaGluProGluValProGlnValGluThr 724  
 DB 2275 CCAACAGAGAACCAAGAGAAATCACCAGAGGAATCAGAAAGAACCTCAGGTCGAGACT 2334  
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspSer 744  
 DB 2335 GAAAGGTTGAGAAAGAACTCAGAGAGGCTGAGATTTACTTGGAAAAATAATTTTAC 2394  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 DB 2395 ATTCAAGTCCCAATGCGAAAGAGACTCTCAGAGATTAAATAATAATTTTACTATTGGC 2454  
 QY 765 IleMetAspAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 DB 2455 ACCAGAGCAACAATATCTATTATGGCAGAAAGCTGAAATACTATTGGCTTTTAAAGGAG 2514  
 QY 785 Ser 785  
 DB 2515 AGT 2517  
 RESULT 10  
 ID AAA65736 standard; DNA; 2647 BP.  
 AC AAA65736;  
 DT 21-NOV-2000 (first entry)  
 DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal; ds.  
 OS Streptococcus pneumoniae.  
 PN WO200039299-A2.  
 PD 06-JUL-2000.  
 PF 20-DEC-1999; 99WO-CA001218.  
 PR 23-DEC-1998; 98US-0113800P.  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 WPI; 2000-452397/39.



Db 1839 GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTTCAGGAAATACCTAGGCAAAA 1898  
 QY 595 SerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeu 614  
 Db 1899 GGACGAGAGCTATCTCAACCGCGTGAAGCAGCTAAGAAGGTGCCACTTGATCGTATG 1958  
 QY 615 ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys 634  
 Db 1959 CCTTACATCTTCATATACTAGTAGAGTCGAGTCAAAACGGTAGTTTAATCATCTCATTTAT 2018  
 QY 635 AspHisTyrHisAsnIleIlePheAlaTrpPheAspAspHisThrTyrLysAlaProAsn 654  
 Db 2019 GACCATTAACCAATCAATCAAAATTTGAGTGGTTGACGAAGCGCTTTATGAGCACCTAAG 2078  
 QY 655 GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrValGluHisProAspGlu 674  
 Db 2079 GGGTATACCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGGAACATCCCAACGAA 2138  
 QY 675 ArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAsp 694  
 Db 2139 CGTCGGCATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAGAAGAAACAAAT 2198  
 QY 695 HisSerGluAspProAsnLysAsnPhenylalanylAsp 706  
 Db 2199 GGTCAAGCTGTATACCAATCAACGGAAGAAACCAAGCGAGGAGAAACCTCAGACAGAAAA 2258  
 QY 706 2259 CCTGAGGAAGAAACCTCGAAGAGAGAAACCAAGCGAGGAGAAACCTCAGACAGAAAA 2318  
 QY 707 2319 CCAACAGAGAAACCAAGAGAGAAATCACCAGAGGATCAGAGAACCTCAGTCCGACT 2378  
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 Db 2379 GAAAGGTTGAAGAAACCACTGAGAGAGGCTGAGATTCTTGGAAAAATCCAGGATCCA 2438  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 Db 2439 ATTATCAAGTCCATGCAAGAGACTCTCACAGGATTAATAATTTACTATTGTCG 2498  
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 Db 2499 ACCCAGCACACAACTATATTATGCGAGAGCTGAAAAAATTTGGCTTTATTAAAGAG 2558  
 QY 785 Ser 785  
 Db 2559 AGT 2561  
 RESULT 11  
 ID ABK15103  
 XX ABK15103 standard; DNA; 2647 BP.  
 AC ABK15103;  
 XX  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE DNA encoding Streptococcus pneumoniae BVH-11.  
 XX  
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;  
 KW streptococcal bacterial infection; gene; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FH Location/Qualifiers  
 FT CDS 45..2567  
 FT  
 FT /\*tag= a  
 FT /product= "BVH-11"  
 FT /note= "The gene is flanked by sequences from the vector  
 FT SP64, no information on which is given in the  
 FT specification"  
 XX

PN WO200198334-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 19-JUN-2001; 2001WO-CA000908.  
 XX  
 PR 20-JUN-2000; 2000US-0212683P.  
 XX  
 PA (SHIR-) SHIRE BIOCHEM INC.  
 XX  
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;  
 XX  
 DR WPI; 2002-122272/16.  
 DR P-FSDB; AAU75933.  
 XX  
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing  
 PT polypeptides, useful as vaccine components for treating or preventing  
 PT streptococcal infections such as otitis media, meningitis, and  
 PT bacteremia.  
 XX  
 PS Disclosure; Fig 4; 113pp; English.  
 XX  
 CC The invention describes an isolated polypeptide (I) with 70-90% identity  
 CC to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or  
 CC BVH-11, or chimeric sequences derived from them. A vaccine (II)  
 CC comprising (I) is useful for therapeutic or prophylactic treatment of  
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an  
 CC individual susceptible to these disorders. (II) is also useful for  
 CC therapeutic or prophylactic treatment of any streptococcal bacterial  
 CC infection (e.g., caused by Streptococcus pneumoniae, group A  
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such  
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or  
 CC Staphylococcus aureus) in an individual susceptible to the infection. A  
 CC polynucleotide (III) encoding (I) is useful in DNA immunisation  
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic  
 CC test for S. pneumoniae infection. (III) is useful for designing DNA  
 CC probes for use in detecting the presence of Streptococcus in a biological  
 CC sample suspected of containing the bacteria. The DNA probes may also be  
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for  
 CC diagnosing streptococcal infections. This sequence encodes the  
 CC Streptococcus pneumoniae protein BVH-11, used to create the antigenic  
 CC peptides described in the method of the invention  
 XX  
 SQ Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4.04e-216 Length: 2647  
 Score: 3218.00 Matches: 616  
 Percent Similarity: 82.95% Conservative: 65  
 Best Local Similarity: 75.03% Mismatches: 102  
 Query Match: 77.26% Indels: 38  
 DB: Gaps: 3  
 US-09-765-271-56 (1-796) x ABK15103 (1-2647)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 105 GCTTATGAACACTAGGTTTGCATCAAGCTCAAACTGTAAAGAAAAATAATCGTGTTCCTAT 164  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 165 ATAGATGGAAAAACAAGCAGCGAAAAACGAGAAATTTGACTCCTGATGAGGTAGCAAG 224  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 225 CGTGAAGGAATCAACCGCCGACAAATCGTCATCAAGATTACGATCAGGTTATGTGACC 284  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 285 TCTCATGGAGACCATTTATCATCTACTATAATGCGAAGGTCCCTTATGATGCCATCATCAGT 344  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

345	Db	GAAGAGCTCTCTCANGAAGAGATCCGAANTATTCAGTTGAAGGATTCAGACATTCGTCAATGAA	404
101	Qy	ValIysGlyGlyTyrValIleLysValAspGlyGlyTyrTyrValTyrLeuLysAspAla	120
405	Db	ATCAAGGGTGGTTATGTGTCATTAAGGTAAACGGTAAATACTATGTTTACCTTAAGGATGCA	464
121	Qy	AlaHisAlaAspAsnValArgThrLysGluGluLileAsnArgGlnLysGlnGluHisSer	140
465	Db	GCTCATCGGATATGTCCTACAAAAGAGAAATCAATCGCGAAAAACAAGACATAGT	524
141	Qy	GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln	160
525	Db	CAGCATCGTGAAGGAGGAGACTTCAGCAACAGATGGTCGGTAGCCITTCACGTTCCACAG	584
161	Qy	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
585	Db	GGACGCTACACCAAGATGATGGTTATATCTTCAATGCATCTGATATCATCGAAGATACG	644
181	Qy	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
645	Db	GGCGATGCCATATTCGTTCTCTCATGGAGATCATTAACATTCAATTCCTAAGATGAGTTA	704
201	Qy	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
705	Db	TCACGTACGAGTTGGCTGCTGCAGAAGCCTTCTATCTGCTCGGGAATAATCTGTCAAAT	764
221	Qy	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
765	Db	TTAAGAACCTTATCCCGACAAAAATAGCGATAAACACTCCAAGAACAACTGGGTACCTTCT	824
241	Qy	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
825	Db	GTAGCAATCCAGGAACTACAAATACTTAAACAAGCAACACCAACCACTAACAGTCAA	884
261	Qy	AlaSerGlnSerAsnAspIleAspSerIleuLysGlnLeuTyrIysLeuProLeuSer	280
885	Db	GCAAGTCAAAGTAATGACATTTAGTAGTCTCTTGAAACAGCTCTACAAACTGCCCTTTGAGT	944
281	Qy	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
945	Db	CAACGCCATGAGAAATCTGATGGCGCTTATTTTCAGCCAGCGCAATCACAAGTCCGAAC	1004
301	Qy	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
1005	Db	GCCAGAGGTGATAGCTGTCCCTCATGTTAAACCATTAACCATTTATCCCTTATGAACAAATG	1064
321	Qy	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
1065	Db	TCTGAATTGGAAACCAAGATTCGTATATTATTCCTTCGTATCGTTTCAAACCATTTGG	1124
341	Qy	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
1125	Db	GTACCAGATTCAAGACCAGAAGAACCAAGTCCACAACCGACTCCAGAACCTAGTCCAAGT	1184
361	Qy	ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer	374
1185	Db	CGCAACTCGACCAAAATCTCTCAACCAAGCTCCAAGCAATCCAAATGTAT-----GAGAAA	1238
375	Qy	LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle	394
1239	Db	TTGFTCAAAGAGCTGTTTCGAAAAGTAGGCGATGGTTATGTTCTTTGAGGAGAAATGGATT	1298
395	Qy	SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys	414
1299	Db	TCTCGTTATATCCACGCCAAAGAACTTTTCAGCAAGAAACAGCAGCAGCATTTGATAGCAAA	1358
415	Qy	LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro	434
1359	Db	CTGGCCACAGCAAGAACTTTATCTCATAGGTAGGAGCTAAGAAAACTGACCTCCCATCT	1418
435	Qy	ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu	454
1419	Db	AGTGATTCGAGAAATTTTACAATAGCGCTTATCACTTACTAGCAAGAAATTCACCAAGTTTA	1478

QY	455	Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeu	474
DB	1479	CTTGATAATAAAGGTGACAAAGTTGATTTTGGAGCGTTTGGATAACCTGTTGGACGACTC	1538
QY	475	AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro	494
DB	1539	AAGGATGCTCAAGTGATTAAGTCAAGTTAGTGGATGATATTCTTGCTTCTTAGCTCCG	1598
QY	495	IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu	514
DB	1599	ATTGCTCATCCAGAAAGCTTTAGAGAAACCAAAATCGCCAAATTACCTACACTGATGAG	1658
QY	515	ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu	534
DB	1659	ATTCAGATGCGCAAGTTGGCAGGCGAGTACCAACAGACAGCGTTATATCTTTGATCCT	1718
QY	535	HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis	554
DB	1719	CGTGATATAACCACTGATGAGGGGATGCTATGTAACTCCATATAGACCCATAGCCAC	1778
QY	555	TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThrLys	574
DB	1779	TGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCAGCGGCTTATGCTAAA	1838
QY	575	GluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAsp	594
DB	1839	GAGAAAGTTTGACCCCTCTCTTCGACAGACCATCAGGATTCAGGAAATACCTAGGCA	1898
QY	595	SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu	614
DB	1899	GGAGCAGAAAGCTATCTCAACCGCGTGAAGACAGCTAAGAAAGTGCACACTTGATCGTATG	1958
QY	615	ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys	634
DB	1959	CTTACCAATCTTCAATATACTGTAGAAGTCAAAACCGTAGTTTAATCATCATCTCATAT	2018
QY	635	AspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsn	654
DB	2019	GACCATTACCATAACATCAAAATTTGAGTGGTTTCAGCAAGCGCTTTATGAGGCACCTAAG	2078
QY	655	GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu	674
DB	2079	GGGTACTCTGTAGGATCTTTTGGCGACCTGTCAAGTACTATGTCGAACATCCAAACGAA	2138
QY	675	ArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAsp	694
DB	2139	CGTCGCGATTCAGATAATGGTTTGGTAAACGTAGCGACCATGTTCAAAGAAACAAAAAT	2198
QY	695	HisSerGluAspProAsnLysAsnPheLysAlaAsp	706
DB	2199	GGTCAAGCTGATACCAATCAAAACGGAACCAAGCGAGGAGAAACCTCAGACAGAAAAA	2258
QY	706	-----	706
DB	2259	CCTGAGGAAGAAACCCCTCGAGAGAGAGAAACCAAAACCGAGAAACACAGAGTCTCCAAA	2318
QY	707	-----GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr	724
DB	2319	CCAAACAGAGGAACCCAGAGAGAGAAATCCACAGAGGAATCAGAAAGAACCTCAGGTGAGACT	2378
QY	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer	744
DB	2379	GAATAAGGTTGAAGAAAAAATCCAGAGAGGCTGAAGATTACTTTGAAAAAATCCAGAGTCCA	2438
QY	745	SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln	764
DB	2439	ATTATCAAGTCCATGCCAAGAGACTCTCCAGAGATTAAAAATAATTTACTATTTGGC	2498
QY	765	IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly	784
DB	2499	ACCCAGGACAAATATCTATTATGTCAGAGCTGAAACATATTGGCTTTTATAAGAG	2558

QY 785 Ser 785  
 Db 2559 AGT 2561  
 RESULT 12  
 ID AAA08557 standard; DNA; 2478 BP.  
 XX AAA08557;  
 XX 19-JUL-2000 (first entry)  
 XX S. pneumoniae 92 kDa human C3-degrading protein coding sequence.  
 XX Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;  
 XX inhibitor; inflammation; organ rejection; xenotransplantation; ss.  
 XX Streptococcus pneumoniae.  
 XX WO200017370-A1.  
 XX 30-MAR-2000.  
 XX 24-SEP-1999; 99WO-US022362.  
 XX 24-SEP-1998; 98US-0101736P.  
 XX 31-MAR-1999; 99US-00283094.  
 XX (MINU ) UNIV MINNESOTA.  
 XX (AMCY ) AMERICAN CYANAMID CO.  
 XX Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;  
 XX WPI; 2000-283594/24.  
 XX P-PSDB; AAY91939.  
 XX Isolated polypeptide is used to stimulate immune system and immunize or  
 XX treat a mammalian subject against Streptococcus pneumoniae infection or  
 XX colonization.  
 XX Claim 1; Page 55-57; 63pp; English.  
 XX The present sequence, isolated from Streptococcus pneumoniae, encodes a  
 XX human C3-degrading protein (see AAY91939) of about 92 kDa. This sequence  
 XX may encompass a smaller 20 kDa polypeptide coding sequence (AAA08556),  
 XX also having human C3-degrading activity. The DNA sequences can be used  
 XX for producing an immune response to Streptococcus pneumoniae in a mammal.  
 XX Antibodies against the proteins can be used to inhibit S. pneumoniae-  
 XX mediated C3 degradation. C3-mediated inflammation and rejection in  
 XX xenotransplantation can be inhibited by expressing the nucleic acid  
 XX sequences on the surface of an organ of an animal. In particular, the  
 XX polypeptides are useful for stimulating the immune system and are  
 XX effective to immunize or treat a mammalian subject against Streptococcus  
 XX pneumoniae infection or colonization  
 XX SQ Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 1.09e-187 Length: 2478  
 Score: 2812.00 Matches: 552  
 Percent Similarity: 77.10% Conservative: 81  
 Best Local Similarity: 67.24% Mismatches: 136  
 Query Match: 67.52% Indels: 52  
 DB: 3 Gaps: 8  
 US-09-765-271-56 (1-796) x AAA08557 (1-2478)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 Db 61 TCCTATGAACCTTGGTCTACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCAGGTTTCT 120  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

Db 121 TATATAGATGGTGATCAGGCTGGTCAAAAGCGCAGAAACCTTGACACAGATGAAGTCACT 180  
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db 181 AAGAGGGAGGGGATCAACGCCGGAACAATCGTCAATCAAGATTACGGATCAAGGTTATGTG 240  
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 Db 241 ACCTCTCATGGAGACCAATTATCTACTATAATGCAAGGTCCCTCTATGATGCCATCATC 300  
 QY 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db 301 AGTGAAGAGCTCTCATGAAGATCCGAATATCGTTGAAGATTACAGCATTTGCAAT 360  
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLeuAsp 119  
 Db 361 GAAATCAAGGTGGTTATGTTATCAAGGTAGTGAATAATCTATGTTTACCTTAAGGAT 420  
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluLuleAsnArgGlnLysGlnGluHis 139  
 Db 421 GCAGCTCATGGGATAATATTCGACAAAAGAGATTAAACGTCAGAAGCAGAAACAC 480  
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db 481 AGTCATAATCACGGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGACGCCAGGCC 534  
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db 535 CAAGGACGCTATACACGGATGATGTTATATCTTCAATCATCTGATATCATTTAGGAC 594  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db 595 ACGGGTATGCTTATATCGTTCTCGCGGACCAATTACCATTCATTCCTCAAGAATGAG 654  
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 655 TTATCAGTACGAGTTAGTCTGTGAGAACGCTTATGGAAATGGG----- 699  
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 Db 700 -----AAGCAGGGATCTCGCTCTTCTTCTCAAGTCTCTAGTTATAATGCA 741  
 QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 742 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCTCACTCCAACTTATCAT 801  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db 802 CAA--AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 858  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db 859 TCAGAACGGCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCACAAGTCGA 918  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db 919 ACGCCAGAGGTAGTCTGCTCCCTCAATGTAACCATTTACCATTTATCCCTTATGAACAA 978  
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 Db 979 ATGCTCTGAATTGAAACAAAGATTGCTCGTATATTTCCTTCTGTTATGCTTCAACCAT 1038  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1039 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCACATCGACTCCGGAACCTAGTCCA 1098  
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 Db 1099 AGTCGCAACCTCGACCAAAATCTTCAACAGCTCCAAAGCAATCCAAATTTGAT-----GAG 1152  
 QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
 Db 1153 AATTTGGTCAAGAGACTCTTCGAAAAGTAGGGGATGTTATGTTCTTTGAGGAGATGGA 1212

QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 DB 1213 GTTTCGTGTTATATCCCAAGGATCTTTCAGCAGAAACAGCAGCAGGAGCATGATGAC 1272  
 QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 DB 1273 AAATGGCCAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAGAGAACTGACCTCCCA 1332  
 QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 DB 1333 TCTAGTGCAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1392  
 QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerPheGlnAlaLeuAspLysLeuLeuGluArg 473  
 DB 1393 TTACTTGTATAATAAGGTCGACAAAGTGTATTTGAGGCTTTGGATTAACCTGTGTGAACGA 1452  
 QY 474 LeuAsnAspGluSerThrAsnLysLysLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
 DB 1453 CTCAGGATGTCCTCAAGTGATAAGTCAAGTTAGTGATGATATTTCTTGCCTTCTAGCT 1512  
 QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
 DB 1513 CCGATTCTGCATCCAGAACGTTTAGCAAAACCAATGGCGCAATTTACCTACACATGATGAT 1572  
 QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp 533  
 DB 1573 GAGATTCAAGTAGCCCAAGTTGGCAGCAAGTACCAACAGAACGCTTATATCTTTGAT 1632  
 QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 DB 1633 CCTCTGATATAACCAAGTATGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 1692  
 QY 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
 DB 1693 CACTCGGATTAATAAGATAGTTTGTGAAGCTGAGAGCGCGCAGCCGAGCTTATGCT 1752  
 QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
 DB 1753 AAAGAGAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGAAATACTGAGGCA 1812  
 QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613  
 DB 1813 AAAGAGCAGCAAGCTATCTACAACCGCTGAAAGCAGCTAAGAGGTCGCACTTGATCT 1872  
 QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
 DB 1873 ATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAACGGTAGTTTATCATCATCTCAT 1932  
 QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
 DB 1933 TATGACCATACCAATCAACATCAATTTGATGTTGAGCAAGGCTTTATGAGGCACT 1992  
 QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
 DB 1993 AAGGGTATACCTTTGAGGATCTTTGGCGACTGTCAAGTACTGTGCAAGATCCAAAC 2052  
 QY 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
 DB 2053 GAACGTCGCATTCAGATAATAGTTTGTGTAACGCTAGGACCATGTTCAAGAACAA 2112  
 QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGlu----- 708  
 DB 2113 AATGGTCAAGCTGATACCATCAACGGGAAACCAACGAGGAGGAGAAACCTCAGACAGAA 2172  
 QY 709 ---ProValGluGluThr----- 713  
 DB 2173 AAACCTGAGGAAGAAACCCCTCGAGAAAGAAACCCGAAAGCGAGAAACACGAGTCTCCA 2232  
 QY 714 -----ProAlaGluProGluValProGlnValGluThr 724  
 DB 2233 AAACCAACAGAGAACCCAGAGAGTAATCACCAGAGGAATCAGAGAACCTCAGGTCGAGACT 2292

QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 DB 2293 GAAAGGTTGAAGAAAAAAGCTGAGAGAGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2352  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 DB 2353 ATTATCAAGTCCAAATGCCAAGAGACTCTCACAGGATTAAAAAATAATTTACTATTGGC 2412  
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly 784  
 DB 2413 ACCCAGACACAATACTATTATGCGAGAGCTGAAAAACTATTGGCTTTATTAAAGGAG 2472  
 QY 785 Ser 785  
 DB 2473 AGT 2475  
 RESULT 13  
 ID ABX06885  
 AC ABX06885; standard; DNA; 2457 BP.  
 DT 27-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain coding region #1173.  
 KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007659.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI: 2003-040579/03.  
 DR P-PSDB; ABU01597.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 6; SEQ ID NO 2345; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a Streptococcus nucleic acid sequence where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein







Db 595 ACGGGTGATGCTTATATCGTTCTCCACGGCGACCATTAACCATTCCTCAAGAAATGAG 654  
 Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 655 TTATCAGCTACGAGTTAGCTGCTGCAAGCCATTTCGAATGGG----- 699  
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 Db 700 -----AAGCAGGATCTCGTCTCTCTCAAGTTCTAGTTATAATGCA 741  
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 742 AATCCAGCTCAACCAAGATTCTCAGAGAACACCAATCTGACTGCACTCCCACTTATCAT 801  
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db 802 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 858  
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db 859 TCAGACGCCATGTGGAATCTGATGGCTTATTTTCGACCAGCGCAAAATCACAGTCGA 918  
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db 919 ACCGCCAGAGGTGTAGCTGCTCCCTCATGGTAAACCATTAACCACTTTATCCCTTATGAACAA 978  
 Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 Db 979 ATGCTGAATGTGNAACCAATTTGCTGTATTTATCCCTTCGTTATCGTTTCAAAACCAT 1038  
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1039 TGGGTACCATGATTCAAGACCAAGAACCAAGTCCCAATCGACTCGGAACCTAGTCCA 1098  
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 Db 1099 AGTCGGCAACCTCGCAAAATCTCAACAGCTCCAAAGCAATCAATGTAT-----GAG 1152  
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyValPheGluGluLysGly 393  
 Db 1153 AAATGGTCAAGAGCTGTTCGAAAGTAGGCGATGTTATGTTCTTGGAGGAATGGA 1212  
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 Db 1213 GTTCTCTGTTATATCCAGCCAAAGATCTTTCAGCAGAAACAGCAGCAGCATTTGATAGC 1272  
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 Db 1273 AAATCGGCCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTGACCTCCCA 1332  
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 Db 1333 TCTAGTGATCGAATTTTACATAGGCTTATGACTTACTAGCAAGATTCACCAAGAT 1392  
 Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
 Db 1393 TTACTTTGATAATAAAGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTTGGAAACA 1452  
 Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuAlaPheLeuAla 493  
 Db 1453 CTAAGGATGTCCTCAAGTAAAGTCAAGTTAGTGGATGATATCTTGGCTTTCTTAGCT 1512  
 Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
 Db 1513 CCGATTCTCATCCAGACGTTTAGGNAACCAATGCGCAATACCTACATCTGATGAT 1572  
 Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
 Db 1573 GAGATTCAAGTACCAAGTTGGCAGGCAAGTACACAAACAGACGGTATATCTTTGAT 1632  
 Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 Db 1633 CTTCTGATATATACCAAGTGTAGGGGGATGCGCTATGTAACTCCACATATGACCCATAGC 1692

Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
 Db 1693 CACTGGAATTAATAAAGATAGTTTCTGTAAGCTGAGAGCGGAGCCAGCTTATGCT 1752  
 Qy 574 LysGluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593  
 Db 1753 AAAGAGAAAGTTTGACCCCTCTTCAGCAGACCATCAGGATTGAGAAATACTGAGGCA 1812  
 Qy 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613  
 Db 1813 AAAGGAGCAGAAAGCTATCTCAACCGGTGAAGCAGCTAAGAAGGTGCCACTTATGCT 1872  
 Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
 Db 1873 ATGCTTTACCAATCTTCAATATACTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1932  
 Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
 Db 1933 TATGACCATTTACCATAACATCAATTTGAGTGGTTTGACGAGGCCCTTTATGAGGCACCT 1992  
 Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
 Db 1993 AAGGGGTATCTCTTGGAGTCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 2052  
 Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
 Db 2053 GAACGTCGCAATTCAGATAATGTTTGGTAAACGCTAGCGACCATGTGTTCAAGAAACAAA 2112  
 Qy 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708  
 Db 2113 AATGGTCAAGCTGATACCAATCAACCGGAAACCAACGAGGAGGAGAACCTCAGACAGAA 2172  
 Qy 709 ---ProValGluGluThr----- 713  
 Db 2173 AAACCTGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGCGAGAAACCGAGAGTCTCCA 2232  
 Qy 714 -----ProAlaGluProGluValProGlnValGluThr 724  
 Db 2233 AAACCAACAGAGAAACAGAGAATCACCAGAGGAATCAGAGAACCCTCAGTCGAGACT 2292  
 Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 Db 2293 GAAAGGTTGAAGAAACCTGAGAGAGCTGAAGATTCTTGAAAAATTCAGGATCCA 2352  
 Qy 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 Db 2353 ATTATCAAGTCCAATGCCAAAGAGACTCTCACAGGATTAATAAATAATTTACTATTGGC 2412  
 Qy 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 Db 2413 ACCCAGACAAACAATATCTATTATGGCAGAGCTGAAAAACTATTGGCTTTATTAAAGGAG 2472  
 Qy 785 Ser 785  
 Db 2473 AGT 2475  
 RESULT 15  
 AAA47605  
 ID AAA47605 standard; DNA; 2531 BP.  
 XX  
 AC AAA47605;  
 XX  
 DT 20-OCT-2000 (first entry)  
 XX  
 DE Recombinant variant of Sp36 gene (Sp36B) of *S. pneumoniae*.  
 XX  
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;  
 KW histidine triad residue; Sp36; antibody; otitis media;  
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;  
 XX meningitis; lobar pneumonia; ds.  
 OS Streptococcus pneumoniae.



Db	1390	CGCAAGTGTGATTTTGAGGCTTTTCGATAA	CTGTGGAACGACTCAAGGATGCTCAAGT	1449
Qy	480	AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu	499	
Db	1450	GATAAAGCTCAAGTTAGTGGAGATATTCCTGCCCTCTTAGCTCCGATTCGTCA	1509	
Qy	500	ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln	519	
Db	1510	CGTTTAGGAAAAACCAATGCGCAATACCTACACTGATGATGAGATTCAAGTAGCC	1569	
Qy	520	LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSer	539	
Db	1570	TTGCAGCGCAGGTACACAGCAGAGACGGTTATATCTTTGATCCTCGTGATATAAC	1629	
Qy	540	AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp	559	
Db	1630	GATGAGGGGATGCTATGTAACTCCACATATGACCCATGACCCCTGGATTA	1689	
Qy	560	SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu	579	
Db	1690	AGTTTGTCTGAAGCTGAGAGCGGCAGCCAGGCTTATGCTRAAGAGAAAGTTTG	1749	
Qy	580	ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAla	599	
Db	1750	CTTCCTTCGACACCATCAGGATTCAGGAAATACTGAGGCAAAAGCAGCAGAG	1809	
Qy	600	TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu	619	
Db	1810	TACAACCGMTGAAGCAGCTAAGAGAGTGCACCTGATCGTATGCTTACAATCTTCA	1869	
Qy	620	HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn	639	
Db	1870	TATACTGTAGAAGCTCAAAACCGTAGTTTAAATCATACCTCATTATGACCATTA	1929	
Qy	640	IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu	659	
Db	1930	ATCAAAATTCAGTGGTTTGACGAAGCGCTTATGAGGCACTTAAGGGGTATAC	1989	
Qy	660	AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn	679	
Db	1990	GATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCGCA	2049	
Qy	680	AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspPro	699	
Db	2050	AAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAATAATGGTCA	2109	
Qy	700	AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr	713	
Db	2110	AATCAACGGMAAAACCAAGCGAGGAGAACTTCAGACAGAAAACTCAGGAAGA	2169	
Qy	713	-----	713	
Db	2170	CCTCGAAGAGAGAAACCGCAAGCGAGAAACACAGAGTCTCCAAAACCAACAG	2229	
Qy	714	-----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln	730	
Db	2230	GAGAATATCCACAGAGAAATCAGAGAAGCACTCAGGTCGAGACTGAAAAAGG	2289	
Qy	731	LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla	750	
Db	2290	CTGAGAGAGCTGAGATTTACTTTGGAATAATCCAGGATCCCAATATATCAAG	2349	
Qy	751	ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer	770	
Db	2350	AAAGAGACTCTCACAGATTAAAAATAATTTACTATTTGGCCACCAGCAGCAAC	2409	
Qy	771	IleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySer	785	
Db	2410	ATTATGGCAGAGCTGAAAAACTATTGGCTTTTATTAAAGCAGAT	2454	

RESULT 16  
ABX06705

ABX06705 standard; DNA; 2517 BP.

ABX06705;

27-OCT-2003 (revised)

11-FEB-2003 (first entry)

S. pneumoniae type 4 strain coding region #993.

Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media; ear infection; antiinflammatory; antibacterial; immunostimulant; auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae; type 4 strain.

WO200277021-A2.

03-OCT-2002.

27-MAR-2002; 2002WO-IB002163.

27-MAR-2001; 2001GB-00007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Masignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

P-PSDB; AB001418.

New proteins and nucleic acid molecules from Streptococcus pneumoniae, useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or ear infection.

Claim 6; SEQ ID NO 1985; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as AB56454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the target sequence and the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2489 identified coding region from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 27-OCT-2003 to standardise OS field)

Alignment Scores:

2517

2410 ATTATGCCAGAGCTGAAAACTATTGGCTTTATTAAAGGAGAGT 2454

RESULT 16  
ABX06705

Score: 2772.00 Matches: 541  
 Percent Similarity: 75.78% Conservative: 91  
 Best Local Similarity: 64.87% Mismatches: 142  
 Query Match: 66.55% Indels: 60  
 DB: 7 Gaps: 8

US-09-765-271-56 (1-796) x ABX06705 (1-2517)

Qy 1 SerTyrGluLeuGlyLeuTyrGluAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 Db TCCTATGAACCTGGTCGTACCAAGCTGGTCAGGTTAAGAAAGAGCTCTAATCGAGTTTCT 120

Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 Db TATATAGATGGTGTAGCTGGTCAAAAGGCGAGAAACTTGCACACCAAGATGAAGTCAGT 180

Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db AAGAGGAGGGAGTCAACGCCGAACAATCGTCATCAAGATTACGATCAAGTTATGTG 240

Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 Db ACCTCTCATGAGACCAATTATCACTACTATATATGCAAGGTCCTTATGATGCCATCATC 300

Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db AGTGAAGAGCTCCTCATGAAGATCCGAATATCACTTGAAGTTCAGACATGTGCAAT 360

Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrIleLysAsp 119  
 Db GAAATCAAGGTTGGTTATGTTATCAAGGTAGATGCAAAATACTATGTTTACCTTAAGGAT 420

Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 Db GCAGCTCATCGCGATAATATCGCAAAAGAGAGATTAAACGTCAAGAGCAGAGAACAC 480

Qy 140 SerGlnHisArgGluGlyGlyTyrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db AGTCATAATACGGGGTGTTCT-----AACGATCAAGCAGTAGTTGCACCCAGAGCC 534

Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db CAAGGACGCTATACAAAGCATGATGTTATATCTTCAATGCATCTCATATCATTTGAGGAC 594

Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db ACGGGTGATGCTTATATCTTCTCACGGCGACCATACCATACATTCCTAAGAAATGAG 654

Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db TTATCAGCTAGCGAGTTAGCTGCTGCAAGAGCTATTTGGAATGGG----- 699

Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 Db -----AAGCAGGAGTCTCGTCCTCTCTCAAGTTCTAGTTTATATATGCA 741

Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db AATCCAGCTCAACCAAGATGTTCAGAGAACCAATCTGACTGTCTCACTCAACTATCAT 801

Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTAA 858

Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProLadnIleThrSerArg 299  
 Db TCAGAACGCCCATGTGGAATCTGATGGCTTATTTCCAGCCAGCGCAAAATCAAGTCGA 918

Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db ACCGCCAGAGGTAGTGTCCCTCAGTGGTAACCATTCACACTTATCCCTTATGACAA 978

Qy 320 MetSerGluLeuGluGluAlaAlaArgIleIleProLeuArgTyrArgSerAsnHis 339

Db 979 ATGTCGTAATTGGAAAAACGAATTGCTCGTATTATTCCTCTCGTTATCGTTCAAAACCAT 1038  
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db TGGGTACAGATTCAAGACCAACCAACCAAGTCCACANTCGATCCGGAACCTAGTCCCA 1038

Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 Db AGTCGCAACCTGACCAATCTCAACAGCTCCAAAGCAATCAAAATTTGAT-----GAG 1152

Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
 Db AAATTGGTCAAAAGAGCTGTTGCAAAAGTAGTGGCATGTTATGTTCTTTGAGGAGAATGGA 1212

Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 Db GTTTCTGTTATATCCCAAGCAAGATCTTTTCAGAGAAACAGCAGCAGGAGGATTTGATAGC 1272

Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 Db AACTGGCCAGCAGGAAAGTTTATCTCAAGCTAGGAGCTAAGAAACTGACCTCCCA 1332

Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 Db TCTAGTCATCAGAAATTTTACATTAAGCTTTATGACTTACTAGCAAGAAATTCACCAAGAT 1392

Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg 473  
 Db TTACTTGTATTAATAAGGTCGCAAGATTGATTTTGGGGCTTTGGATAACTGTTGGAACGA 1452

Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
 Db CTCAGGATGTCCTCAAGTGATAAAGTCAAGTTAGTGATGATATTTCTTGCCTTCTTAGCT 1512

Qy 494 ProfileThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
 Db CCGATTCTCATCCAGAACGTTTAGGAAACCAATTCGCAAAATTTACCTTACACTGATGAT 1572

Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
 Db GAGATTCAAGTAGCCAGTTGGCAGCAGTACACAAAGAGAGCGTTATATCTTTGAT 1632

Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 Db CCTCGTATATAACCAAGTATGAGGGGATGCTATGTAATCCACATATGATGATGATG 1692

Qy 554 HistTrpIleGlyLysAspSerLeuSerLysGluLysValAlaAlaGlnAlaTyrThr 573  
 Db CACTCGATTAAAAAAGTAGTTTGTCTGAAGCTGAGAGCGCGAGCCCGAGGCTTATGCT 1752

Qy 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
 Db AAGAGAGAGGTTTGACCTCTTCGACAGACCATCAGGATTCAGGAAATACTGAGGCA 1812

Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613  
 Db AAAGGAGCAGAGCTATCTACAAACCGCTGAAAGCAGCTAAGAAAGTGCCTTGTATCGT 1872

Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnGluLeuIleProHis 633  
 Db ATGCGCTTACAATCTTCAATATATCTGAGAGGTCAGAAACCGTAGTTTATATCATCTCAT 1932

Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
 Db TATGACCATTACATTAACATCAAAATTTGAGTGTGTTGACGAGCCCTTTTATGAGCACCT 1992

Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
 Db AAGGGGTATACTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCAAA 2052

Qy 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys--- 692



Db 2053 GAACGTCGCGATTTCAGATAATGTTTGGTAACGCTAGCAGCATGTTTCGTAATAAATAG 2112

QY 693 -----lysAspHisSerGlu----- 697

Db 2113 GTAGACCAAGACAGTAACCTGTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACT 2172

QY 697 ----- 697

Db 2173 CACCTGAATCTGATGAAAGAGAAATCACGCTGGTTAAATCTTTCAGCAGATAAATCTT 2232

QY 698 ---AspProAsnLysAsnPhelLysAlaAspGluProValGluGluThrProAlaGlu 716

Db 2233 TATAAACAAGCACTGATACGGAAGACAGACAGGAGAAAGCTGAAGATACACACAGATGAG 2292

QY 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736

Db 2293 GCTGAATCTTCAAGTAGAAGATCTGTTATTAACTGCTAAGATAGCAGATGCGGAGGCC 2352

QY 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756

Db 2353 TTGCTAGAAAAGTAACAGATCTTAGTATTAGACAAAATGCTATGGAGACATTCAGTGGT 2412

QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMecAspAsnAsnSerIleMetAlaGluAlaGlu 776

Db 2413 CTAAAAAGTAGTCTCTCTCGGAACGAAAGATAATAACACTATTTCAGCAGAAAGTAGAT 2472

QY 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal 790

Db 2473 AGTCTCTGGCTTGTGTTAAAGAAAGTCAACCGGCTCCTATA 2514

RESULT 17

AAAA47602

ID AAA47602 standard; DNA; 2531 BP.

XX AAA47602;

XX 20-OCT-2000 (first entry)

XX Recombinant variant of Sp36 gene (Sp36D) of S. pneumoniae.

XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;

KW histidine triad residue; Sp36; antibody; otitis media;

KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;

XX meningitis; lobar pneumonia; ds.

XX Streptococcus pneumoniae.

XX Key Location/Qualifiers

FT 1. .2517

CDS /\*tag= a

FT /product= "Sp36D polypeptide"

FT

XX WO200037105-A2.

XX 29-JUN-2000.

XX 21-DEC-1999; 99WO-US030390.

XX 21-DEC-1998; 98US-0113048P.

XX (MEDI-) IMMUNE INC.

XX Johnson LS, Koenig S, Adamou JE;

XX WPI; 2000-452129/39.

DR P-PSDB; AAB01466.

XX Vaccine useful for prophylaxis and treatment of pneumococcal infections

PT such as otitis media, nasopharyngeal and bronchial infections, comprises

PT Streptococcus pneumoniae proteins.

XX Disclosure; Page 57-58; 70pp; English.

XX

CC Although a number of proteins have been suggested as being involved in

CC the pathogenicity of Streptococcus pneumoniae, there still remains a need

CC to identify polypeptides having epitopes in common from various strains

CC of S. pneumoniae in order to utilize such polypeptides in vaccines to

CC protect against a wide variety of S. pneumoniae. New vaccine compositions

CC are described which comprise a Streptococcus pneumoniae polypeptide (or

CC fragments) of 80 - 680 amino acids in length that comprise at least one

CC histidine triad residue (HxxHxxH) or a coiled-coil region, or an antibody

CC directed against these features. The vaccine is useful in protecting

CC against infection by Streptococcus pneumoniae. The vaccine composition

CC comprising antibodies to is useful for passive immunization for treating

CC Pneumococcal infections which includes otitis media, nasopharyngeal and

CC bronchial infections

XX

Sequence 2531 BP; 836 A; 511 C; 547 G; 637 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 71e-185 Length: 2531

Score: 2772.00 Matches: 541

Percent Similarity: 75.78% Conservative: 91

Best Local Similarity: 64.87% Mismatches: 142

Query Match: 66.55% Indels: 60

DB: 3 Gaps: 8

US-09-765-271-56 (1-796) x AAA47602 (1-2531)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19

Db 61 TCCTATGAACCTGCTGCTACCAAGCTGCTGAGTTAAGAAAGAGTCTAATCGAGTTTCT 120

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

Db 121 TATATAGATGCTGATCAGGCTGCTCAAAAGCGCAGAAACTTGACACAGATGAAGTCAGT 180

QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59

Db 181 AAGAGGGAGGGGATCAACGCCGACAAATCTCATCAAGATTACGATCAAGGTTAATGTG 240

QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIle 79

Db 241 ACCTCTCATGGAGACCATTAATTAATAATGCAAGCTCCCTTATGATGCCATCATC 300

QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99

Db 301 AGTGAAGAGCTCCTCATGAAGATCCGAATTAATCAGTTGAAGGATTCAGACATGTCAT 360

QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

Db 361 GAAATCAAGGCTGTTATGTTATCAAGGTAGATGGAATAATATCTATGTTTACCTTAAGAT 420

QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139

Db 421 GCAGCTCATGCGGATAATATTCGACAAAAGAGAGATTAAACGTCAGAGCAGGAACAC 480

QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159

Db 481 AGTCATAATCACGGGGTGGTCTCT-----AACGATCAACAGCAGTAGTTGCGAGCAGAGCC 534

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179

Db 535 CAAGGACGCTATACACGCGATGATGTTTATCTTCAATGTCATCTGATATCATTTAGGAC 594

QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199

Db 595 ACGGGTGATGCTTATATCTCTCCACGGCGACCATTAACCATTTACATTCCTAAGAATGAG 654

QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219

Db 655 TTATCAGTAGCGAGTAGTCTGCTGCAAGAGCCTATTGGAATGGG----- 699

QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239

Db 700 -----AAGCAGGGATCTCGTCTCTTCTTCAAGTTCCTAGTTATATAATGCA 741



Qy	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
Db	742	AATCCAGCTCAACCAAGATGTCAGAGAACCAATCTGACTGTCTCACTCCAACCTATCAT	801
Qy	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279
Db	802	CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTTA	858
Qy	280	SerClnAArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
Db	859	TCGAAGCCGATCTGGAAATCTGATGGCCTTATTTTCGCCAGCGCAATACAAAGTCGA	918
Qy	300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319
Db	919	ACGCCACAGAGGTGTAGCTGCCCTCATGTGTAAACCATTCACACTTATCCCTTATGAACAA	978
Qy	320	MetSerGlnLeuGluAurGlnIleAlaAArgIleIleProLeuArgTyrArgSerIleHis	339
Db	979	ATGTCTCAATTTGGAATAACCAATTTGCTGCTATTATTCCTTCGTATCGTTCAAACCAT	1038
Qy	340	TripValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
Db	1039	TGGTACCAGATTCAGACCAGAACCAACCAAGTCCCAATCGAATCCGAAACCTAGTCCA	1098
Qy	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
Db	1099	AGTCGGCAACCTGCACCAAAATCTCTCAACCAAGCTCCCAAGCAATCAAAATTGAT-----GAG	1152
Qy	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
Db	1153	AAATTTGCTCAAGAAGCTGTTTCGAAAGTAGGCGATGTTATGCTCTTGAGGAGATGGA	1212
Qy	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
Db	1213	GTCTTCTCGTTATATCCCAAGCAAGATCTTTACAGCAAAACAGCAGCAGCATTTGATAGC	1272
Qy	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433
Db	1273	AAACTGGCCAAAGCAGGAAGTTTATCTCAAGTAGAGCTAAGAAAACCTGACCTCCCA	1332
Qy	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453
Db	1333	TCTAGTGATCGAGAATTTTCAATAAGGCTTATGACTTACTAGCAAGATTCACCAAGAT	1392
Qy	454	LeuPhe***AsnLysGlyArgAsnSerAspPheClnAlaLeuAspLysLeuLeuGluArg	473
Db	1393	TTACTTGATAATAAAGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGACCGA	1452
Qy	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla	493
Db	1453	CTCAAGGATGCCCAAGTGATAAAGTCAAGTTAGTGGATGATATCTTGCTTCTTAGCT	1512
Qy	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
Db	1513	CCGATTTCTCATCCAGACAGTTTATAGGAAAAACCAATCGCAAAATACCTACACTGATGAT	1572
Qy	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
Db	1573	GAGATTCAGTATGCCAAGTTGGCAGCGAAGTACCAACAGACGCGTTATATCTTTGAT	1632
Qy	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
Db	1633	CCTCTGTATATAAACAGTGTAGCGGGATGCCCTATGTAACTCCACATATGACCCCATAGC	1692
Qy	554	HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573
Db	1693	CACTGGATTAATAAAGAATGTTTGTCTGAAGCTGAGAGCGCGCAGCCGAGGCTTATGCT	1752
Qy	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
Db	1753	AAAGAGAAAGGTTTTCGCCCTCTTCGACAGACCATCAGGATTCAGGAAATCTGAGGCA	1812

## RESULT 18

RESUL 16  
ABS56454 08

Continuation

WP Sequence

WP Fragr

WP  
ABS50

WP  
ABS50

WP ABS50

WP ABS50

WP ABS50

WP ABS50

WP	ABS50
WP	ABS50

WP	ABS50
WP	ADGE

WP	ABS50
WP	ABS50

WP	ABS50
WD	ABS50

WP ABS51

WP ABS51

WP ABS51  
WP ABS51

WP ABS51  
WP ABS51

WP ABS50  
WP ABS50

WF ABS50  
WP ABS50

WP  
ABS5

WP ABS5

WP ABS5

WP	ABS56454_20	2000001	2110000
WP	ABS56454_21	2100001	2162598
Alignment Scores:			
Pred. No.:	6.57e-183	Length:	110000
Score:	2772.00	Matches:	541
Percent Similarity:	75.78%	Conservative:	91
Best Local Similarity:	64.87%	Mismatches:	142
Query Match:	66.55%	Indels:	60
DB:	7	Gaps:	8
US-09-765-271-56 (1-796) x ABS56454_08 (1-110000)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaAargThrVal---	LysGluAsnAsnArgValSer 19
Db	38009	TCTATGAACCTGTGCTGCACCAAGCTGGTCAGGTAAAGAAGAGTGCTAATCGAGTTTCT	38068
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39
Db	38069	TATATAGTGTGATCAGGTGGTCAAAGGCCAGAAAACCTTGACCAGATGAAGTCAGT	38128
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59
Db	38129	AAGAGGAGGGGATCAACGCCGAACAATCGTCATCAAGATTACGGATCAAGGTTATGTG	38188
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyraSngLysValProTyrAspAlaIlelle	79
Db	38189	ACCTCTCATGGAGACCATTAATCAATTACTATAATGGCAAGTCCTTATGATGCCATCATC	38248
QY	80	SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99
Db	38249	AGTGAAGAGCTCTCATGAAGATCCGAATTATCAGTTGGAAGGATTCAGACATTTGTCAAT	38308
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119
Db	38309	GAATCAAGGCTGTTATGTTATCAAGTAGATCGAAAAATACTATGTTTACCTTAAGGAT	38368
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHuHis	139
Db	38369	GCAGCTCATCGGATATATTCGCAAAAAGAAGAGATTAAACGTCAAGACGGAACAC	38428
QY	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159
Db	38429	AGTCATATCACGGGGTGGTCT-----AACGATCAGCAGTAGTTCGACGACAGCG	38482
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp	179
Db	38483	CAAGGACGCTATACACGGATGATGGTTATCTTCAATGCAITCTGATATCATTTGAGGAC	38542
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199
Db	38543	ACGGGTGATGCTTATATCGTTCCTCAGCGGACCATTCACATTACATTCCTAAGAATGAG	38602
QY	200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219
Db	38603	TTATCAGTAGCAGTGTAGTGTCTGCAGAACCTTATGGAAATGGG-----	38647
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239
Db	38648	-----AAGCAGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCA	38689
QY	240	SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
Db	38690	AAATCCAGCTCAACCAAGATTGTCAGAGAACCACAAATCTGACTGTCTCACTCCAATCAT	38749
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279
Db	38750	CAA--AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTTA	38806
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
Db	38807	TCAGAACGCGCATGTGAAATCTGATGCCCTTATTTTCGACCCGCGCAATATCAAGTCGA	38866
QY	300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319
Db	38867	ACGCCAGAGGTAGTGTCTCTCATGTATACCATTACACCTTTATCCCTTATGAACAA	38928
QY	320	MetSerGluLeuGluGlyAlaAlaArgIleIleProLeuArgTyrArgSerAsnHis	339
Db	38927	ATGCTCTGAATTGAAAAACGAATTGCTGTATTTATTCCTCTCGTTATGTTCAAACCAT	38988
QY	340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
Db	38987	TGGGTACCAAGATTCAAGACCAACCAAGTCCCAATGACATCGGAACCTAGTCCA	39046
QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
Db	39047	AGTCCGCAACCTGCACCAATCTCAACAGCTCCNAGCAATCCAAATTGAT-----	39100
QY	374	SerLeuValSerGlnLeuValArgLysValGlyGlyGlyTyrValPheGluGluLysGly	393
Db	39101	AAATTGCTCAAAAGAAGCTGTTCCAAAAGTAGCGATGGTTATGTCTTTGAGGAGAATGGA	39160
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
Db	39161	GTITCTCGTTATATCCAGCCCAAGGATCTTTCAGAGAACACAGCAGCAGGCAATTCATAGC	39220
QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433
Db	39221	AAACTGGCCACGACGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACAGTCACTCCCA	39280
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453
Db	39281	TCTAGTGTATCGAAGATTTTACAAATAAGGCTTATGACTTACTACGAAAGAAATTCACCA	39340
QY	454	LeuPhe**AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg	473
Db	39341	TTACTTGTATATAAAGGTCGACAGTGTATTTTTGAGGCTTTGGATTAACCTGTGGAACGA	39400
QY	474	LeuAsnAspGluSerThrAsnLysGlyLysLeuValAspAspLeuLeuAlaPheLeuAla	493
Db	39401	CTCAAGGATGTCCTCAAGTCATAAAGTCAAGTTAGTGATGATATTTCTTGCCTTCTTAGCT	39460
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
Db	39461	CGATTCGTCTATCCAGAACGTTTAGGAAAAACCAATGCGCAATTACCTACACTGATCAT	39520
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
Db	39521	GAGATTCAAGTAGCCAGTGTGCGGCAAGTACACACAGAAAGCGGTTATATCTTTGAT	39580
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
Db	39581	CCTCGTGATATAACCCAGTGATGAGGGGATGCCATATGTAATCTCCACATATGACCCATAGC	39640
QY	554	HisTrpIleGlyLysAspSerLeuSerAspLysGlnLysValAlaAlaGlnAlaTyrThr	573
Db	39641	CAGTGGATTAAAAAGATAGTTGTCTGAAGCTGAGAGCGGACGCGCCAGCTTATGCT	39700

Db 39841 AAGGGGTATCTCTGAGATCTTTGGCGACTGTGCACTACTATGTCGACATCCAAAC 40000  
 Qy 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys 692  
 Db 40001 GAACGTCCGCAATTCAGATAATGGTTTGGTAAACGCTAGCGACCATGTTCTGTAATAAATAAG 40060  
 Qy 693 -----LysAspHisSerGlu----- 697  
 Db 40061 GTAGACCAAGACAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAAGTGAGCCAACT 40120  
 Qy 697 ----- 697  
 Db 40121 CACCCTGAATCTGATGAAGAAAGAGAAATCACGCTGGTTTAAATCCTTCAGCAGATAATCTT 40180  
 Qy 698 ---AspProAsnLysAsnPhelYsAlaAspGluProValGluGluThrProAlaGlu 716  
 Db 40181 TATAAACCAAGCACTGATACGGAAGACAGACAGAGGAAGAGTGAAGATACACAGATGAG 40240  
 Qy 717 ProGluValProGluValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736  
 Db 40241 GCTGAAATTCCTCAAGTAGAGAAATCTGTTATTAAACGCTTAGATAGCAGATGCGGAGGCC 40300  
 Qy 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756  
 Db 40301 TTGCTAGAAAGAACTAACAGATCTCTAGTATTAGACAAATGCTATGAGACATTTGACTGCT 40360  
 Qy 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776  
 Db 40361 CTAAAGAGTAGTCTTCTTCGGAACGAAAGATAATAACACTATTTCAGCAGAGTAGAT 40420  
 Qy 777 LysLeuLeuAlaLeuLysGlySerAsnProSerVal 790  
 Db 40421 AGTCTCTGGCTTTGTTAAAGAAAGTCAACCGGCTCTTATA 40462

RESULT 19  
 AAA65737  
 ID AAA65737 standard; DNA; 2639 BP.  
 XX  
 AC AAA65737;  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;  
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
 KW otitis media; pneumonia; immunisation; bactericidal; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN WO200039299-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PF 20-DEC-1999; 99WO-CA001218.  
 XX  
 PR 23-DEC-1998; 98US-0113800P.  
 XX  
 PA (BIOC-) BIOCHEM PHARMA INC.  
 XX  
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;  
 XX  
 DR WPI; 2000-452397/39.  
 DR P-PSDB; AAB12720.  
 XX  
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
 PT otis media, bacteraemia and/or pneumonia.  
 XX  
 XX Example 6; Fig 16; 106pp; English.  
 XX  
 CC The present invention describes nucleic acids (I) encoding protein  
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens have

CC bactericidal activity. The nucleic acids, encoding the protein antigens,  
 CC may be used for the recombinant production of the proteins they encode.  
 CC The protein antigens may then be used as vaccines for the prevention and  
 CC treatment of Streptococcal infections in mammals (especially humans)  
 CC which result in, e.g. meningitis, otitis media, bacteraemia and/or  
 CC pneumonia. The present sequence encodes the S. pneumoniae BVH-11-2  
 CC protein antigen  
 XX  
 SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,4e-183 Length: 2639  
 Score: 2750.50 Matches: 538  
 Percent Similarity: 75.30% Conservative: 90  
 Best Local Similarity: 64.51% Mismatches: 145  
 Query Match: 56.04% Indels: 61  
 DB: 3 Gaps: 8  
 US-09-765-271-56 (1-796) x AAA65737 (1-2639)  
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 Db 174 TCCTATGAACTTGGTCGTCCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTCT 233  
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 Db 234 TATATAGATGCTGATCAGGCTGGTCAAAAGGCAGAAAATTTGACACCAGATGAAGTCAGT 293  
 Qy 40 LysArgGluGlyIleAsnAlaGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db 294 AAGAGAGAGGGGATCAACCGCAACAAATTTATCAAGATTACCGATACAGGTTATGTG 353  
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 Db 354 ACCTCTCATGAGACCATTAATCTATATATGCAAGGTTCTTATGATGCCATCATC 413  
 Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db 414 AGTGAAGAACTTCTCATGAAAGATCCGAATTTATCAAGTTGAAGATTCAGACATTTGCAAT 473  
 Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
 Db 474 GAAATCAAGGGTGGCTATGATTAGGTAGCGGAATAATCTATGTTTACCTTAAGAT 533  
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 Db 534 GCGGCCCATCGGACAAATATTCGGACAAAGAGAGATTAAACGTCAGAAGCAGGAACAC 593  
 Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db 594 AGTCATAATCATAACTCA-----AGAGCAGATAATGCTGTGCTGCAGCCAGAGCC 644  
 Qy 160 GlnGlyArgGlyThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db 645 CAAGGACGTTATACACGGATGATGGTATATCTTCAATGCACTCATCATCATGAGGAC 704  
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db 705 ACGGGTGATGCTTATATCTCTCTCACGGCAGACCATTAACATTACATTCTTAAGAATGAG 764  
 Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 765 TTATCAGCTAGCGAGTTAGCTGCTGCAGAAAGCTATTGGAATGGG----- 809  
 Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 Db 810 -----AAGCAGGATCTCGCTCTTCTTCAAGTTCTAGTTATATGCA 851  
 Qy 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 852 AATCCAGTTTCAACCAAGATTGTCAGAGAACCAACATCTGACTGTCTCACTCAACTTATCAT 911  
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279

```

Db 912 CAA---AATCAGGGGAAACATTTCAGCGCTTTTACGTGAATGTATGCTAAACCCCTTA 968
Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 969 TCAGAACGCCATGTAGAACTCGATGGCTTATTTTCGACCCAGCGCAATCACAAGTCGA 1028
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 1029 ACCGCCAGAGGTGAGCTGCTCCCTCATGGTAACCAATACCACTTTATCCCTTATGAACAA 1088
Qy 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db 1089 ATGCTGCAATTGGAAAAAGCAATGCTCGTATTATTCCTCTCGTATCGTTCAAACCAT 1148
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer--- 358
Db 1149 TGGGTACCAGATTCAAGACCAAGAACACCAAGTCACCAATCGATCCGGAACCTAGTCCA 1208
Qy 359 -----ProGlyProGlnProAlaProLeuLeuLysIleAspSerAsnSer 373
Db 1209 AGTCTGCAACTGCACCAATCTCAACCGCTCAAGCAATCCAAATGAT-----GAG 1262
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
Db 1263 AAATTGGTCAAAAGAGCTGTTTCGAAAAGTAGCGCATGCTTATGCTTTGAGGAGAAATGA 1322
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1323 GTTTCCTCGTATATCCCGACGCAAGGATCTTTTCAGCAGAAACAGCAGCAGGATGATAGC 1382
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysLysGluAsnValAla 433
Db 1383 AACTGCGCCAGCAGGAAAGTTTATCTCATAGCTAGAGCTAAGAAAACCTGACCTCCCA 1442
Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1443 TCTAGTATCGAGAATTTTCAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1502
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1503 TTACTTGATATAAAGGTCGACAAAGTTGATTTGAGGTTTGGATAACCTGTTGGAACGA 1562
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493
Db 1563 CTCAGGATGCTCAAGTGAATAAGTCAAGTTAGTGAATGATATCTTGCTTCTTAGCT 1622
Qy 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1623 CCGATTGCTCATCCAGAACGTTTAGGAAAACCAATCGCAAAATTCCTACACTGATGAT 1682
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1683 GAGATTCAAGTAGCCAAAGTTGGCAGGCAAGTACACACAGAAAGCGGTATATCTTTGAT 1742
Qy 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1743 CCTCGTGATATAACCATGATAGGGGGATGCCCTATGTAATCTCACATATAGCCCATAGC 1802
Qy 554 HisTrpIleGlyLysAspSerLeuSerAspLysTyrThrThrSerAspGlyTyrIlePheAsp 573
Db 1803 CACTGGATTAAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCT 1862
Qy 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1863 AAAGAGAAAGGTTTACCCCTCTTCACAGACCCAGGATTCAGGAAATCTAGGCA 1922
Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
Db 1923 AAGGAGCAGCAAGCTATCTAACCCGGTGAAGAGCAGCTAAGAGGTCACCTTGATCGT 1982
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633

```

```

Db 1983 ATGCCTTACAATCTTCAATATACCTGTAGAACTCAAAAACGGTAGTTTAAATACATCAT 2042
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 2043 TATGACCATTAACCAATCAATTCAGTTGGTTTGACGAGGCCCTTTATGAGGCACCT 2102
Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 2103 AAGGGGTATAGTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAC 2162
Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692
Db 2163 GAACGTCCGCATTGAGATATGGTTTGGTAACCGTAGTGACCATGTTCTGTAATAAATAAG 2222
Qy 693 -----LysAspHisSerGlu----- 697
Db 2223 GCAGACCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAGTGAGCCAACT 2282
Qy 697 ----- 697
Db 2283 CACCTGAATCTGATGAAAAAGAGAAATCACGCTGGTTTAAATCTCTCAGCAGATAATCTT 2342
Qy 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716
Db 2343 TATAAACCAAGCACTGATACGGAAGACAGACAGAGAGAAAGCTGAAGATACCACAGATGAG 2402
Qy 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736
Db 2403 GCTGAAATCTCTCAAGTAGAATTCCTGTTATTAACTGAAGTAGCAGATCGGAGGCC 2462
Qy 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaIleAlaThrGluThrLeuAlaGly 756
Db 2463 TTGCTAGAAAAGTAAACAGATCTCTAGTATTAGACAAAATGCTATGAGACATTCAGCTGT 2522
Qy 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
Db 2523 CTAAAAAGTAGTCTCTCTCGGACGAAAGATAATAACACTATTTCAGCAGAAATGAT 2582
Qy 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal 790
Db 2583 AGTCTCTGGCTTTGTTTAAAAAGAAAGTCAACCGGCTCCTATA 2624

RESULT 20
ABK15104
ID ABK15104 standard; DNA; 2639 BP.
XX
AC ABK15104;
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11-2.
XX
BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia;
KW streptococcal bacterial infection; gene; ds; BVH-11-2.
XX
OS Streptococcus pneumoniae.
XX
PH Key Location/Qualifiers
CDS 114..2630
FT /*tag= a
FT /product= "BVH-11-2"
FT /note= "The gene is flanked by sequences from the vector
FT SP64, no information on which is given in the
FT specification"
XX
PN W0200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA000908.
XX
PR 20-JUN-2000; 2000US-0212683P.
XX

```

PA (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

PI WPI; 2002-122272/16.

XX P-PSDB; AAU75934.

DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing

XX polypeptides, useful as vaccine components for treating or preventing

PT streptococcal infections such as otitis media, meningitis, and

PT bacteremia.

XX Example 3; Fig 5; 113pp; English.

PS The invention describes an isolated polypeptide (I) with 70-90% identity

XX to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or

CC BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection. A

CC polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence encodes the

CC Streptococcus pneumonia protein BVH-11-2, used to create the antigenic

CC peptides described in the method of the invention

XX SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,4e-183 Length: 2639

Score: 2750.50 Matches: 538

Percent Similarity: 75.30% Conservative: 90

Best Local Similarity: 64.51% Mismatches: 145

Query Match: 66.04% Indels: 61

DB: 6 Gaps: 8

US-09-765-271-56 (1-796) x ABK15104 (1-2639)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19

DB 174 TCCTATGAACCTTGGTGGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 233

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39

DB 234 TATATAGATGGTGATCAGCTGGTCAAAAGGCGAGAAATTTACACACAGATGAAGTCAGT 293

QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59

DB 294 AAGAGAGGGGATCAACCCGAGACAAATTTGTTATCAAGATTACGGATCAAGTTATGTG 353

QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79

DB 354 ACCTCTCATGGAGACCATATATCAATTAATTAATGCGAAGTCTCTTATGATGCGATCATC 413

QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99

DB 414 AGTGAAGAACTTCTCATGAAGAAATCCGAATTAATCAAGTGAAGATTCAGACATGTGCAAT 473

QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

DB 474 GAAATCAAGGTGGCTATGTGATTAAGTAGACGGAATAATCTATGTTACCTTAAAGAT 533

QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139

DB 1563 CTCAGGATGCTCTCAAGTGATTAAGTCAAGTTAGTGGATGATATCTTCCCTCTTCTAGCT 1622

DB 534 GCGGCCCATGCGGACAAATATTTCGACAAAAGAGAGATTAAACGTCAGAACGAGAACAC 593

QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159

DB 594 AGTCATTAATCAATACTCA-----AGACGATATATGCTGTTGCTGACGACGAGCC 644

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179

DB 645 CAAGACGCTTATCAACGAGATGATGGGTATATCTTCAATGATCTGATATCATTTGAGGAC 704

QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199

DB 705 ACGGTGATGCTTATATGCTTCTCAGCGGACCATTACCATTCATTCCTTAAGAATGAG 764

QY 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219

DB 765 TTATCAGCTAGCAGATGATGCTGCTGAGAACCTTATGGAATGGG----- 809

QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239

DB 810 -----NAGCAGGATCTCGTCTCTTCAAGTCTTAGTTATAATGCA 851

QY 240 SerValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259

DB 852 AATCCAGTTCACCAAGATGTCAGAGAACCAACATCTGACTGTCTCACCACCTTATCAT 911

QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279

DB 912 CAA---AATCAAGGGGAAACATTTCAAGCTTTACGTGAATGTATGCTTAAACCTTA 968

QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299

DB 969 TCAGAACCCATGAGATCTGATGGCTTATTTTCGACCCAGCCAAATCACAAGTCCA 1028

QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319

DB 1029 ACCGCCAGAGGTGATGCTGCTCCCTCATGTAACCATACCTTATCCCTTATGAAACA 1088

QY 320 MetSerGluLeuGluGluAlaArgIleIleProLeuArgTyrArgSerAsnHis 339

DB 1089 ATGCTCTGAATTTGGAAGAAAGATTTGCTGATATTTATCCCTTCGTTATCGTTCAACCAT 1148

QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer--- 358

DB 1149 TGGGTACAGATTCAGACCCAGAACCAAGTCCACAAATCGACTCCGGAACCTAGTCCA 1208

QY 359 -----ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSer 373

DB 1209 AGTCTGCAACCTGCAACCAATCTCAACAGCTCCAGCAATCCCAATTGAT-----GAG 1262

QY 374 SerLeuValSerGlnLeuValArgLysValGlyGlyGlyTyrValPheGluGlyLysGly 393

DB 1263 AAATTTGGTCAAGAGCTGTTGAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGGA 1322

QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413

DB 1323 GTTCTCTGTTATATCCAGCCCAAGGATCTTTCAGCAGAAACAGCAGCAGGCAATTGATAGC 1382

QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433

DB 1383 AAATCGGCCACAGAGAAAGTTTATCTCTAAGCTAGGAGCTAAGAAACTGACCTCCCA 1442

QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453

DB 1443 TCTAGTGTATCGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1502

QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArg 473

DB 1503 TTACTTGTATAATAAAGGTCGCAAGTTGATTTTGAGGTTTGTGATTAACCTGTTGAAACGA 1562

QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla 493

DB 1563 CTCAGGATGCTCTCAAGTGATTAAGTCAAGTTAGTGGATGATATCTTCCCTTCTTAGCT 1622

QY 494 ProfileThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluThrGluAsp 513  
 DB 1623 CCGATTCGTATCCAGACAGCTTTAGGAACCAATGCGCAATTAACCTACACTGAT 1682  
 QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
 DB 1683 GAGATTCAAGTAGCCAGTTCGCGAGGCAAGTACACAGAGAGCGGTATATCTTTGAT 1742  
 QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 DB 1743 CCTCGTATATAACAGTATGAGCGGGATCCCTATGTAACTCCACATATGACCCATAGC 1802  
 QY 554 HistripIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
 DB 1803 CACTGGATTAATAAAGATAGTTTCTGGAAGCTGAGAGCGGAGCCAGCGCTATGCT 1862  
 QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
 DB 1863 AAAGAGAAAGGTTTGACCCCTCTTCGACAGACACCAGGATTTCAGGAAATACTGAGGCA 1922  
 QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613  
 DB 1923 AAGGAGCAGAGCTATCTCAACCGGTGAAGCAGCTAAGAAGGTGCCACTTGTATCGT 1982  
 QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
 DB 1983 ATGCTTACATCTTCATATATCTGTAGAAGTCAAAACCGTAGTTTAATCATACCTCAT 2042  
 QY 634 LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653  
 DB 2043 TATGACCATTAACATACATCAATTTGAGTGGTTTGACGAGGCGCTTTATGAGGCACCT 2102  
 QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
 DB 2103 AAGGGGTATAGTCTGTAGGATCTTTGGCGACTGCTCAAGTACTATGTCGAACATCCAAAC 2162  
 QY 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys--- 692  
 DB 2163 GAACGTCGCGATTCAGATTAATGTTTGGTAAACGCTAGTACCATTGTCGTAATAAATAAG 2222  
 QY 693 -----LysAspHisSerGlu----- 697  
 DB 2223 GCAGACCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGACCAACT 2282  
 QY 697 ----- 697  
 DB 2283 CACCCTGAATCTGATGAAAAGAGAAATCACGCTGGTTTAAATCCTTCAGCAGATAATCTT 2342  
 QY 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716  
 DB 2343 TATAAACCAAGACTGATACGGAAGACAGAGAGGAAGAGCTGAAGATACACAGATGAG 2402  
 QY 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736  
 DB 2403 GCTGAAATCTCTCAAGTAGAGAAATCTGTATTAAACGCTAAGATAGCAGATGCGAGGCC 2462  
 QY 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756  
 DB 2463 TTCTGAAAAGATTAACAGATCTCTAGTATTAGACAAATCTATGGAGACATTTGACTGGT 2522  
 QY 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776  
 DB 2523 CTAAAGTAGTCTTCTCTCGNACCAAGATAATTAACACTATTTCAGCAGAGTAGAT 2582  
 QY 777 LysLeuLeuAlaLeuLysGlySerAsnProSerSerVal 790  
 DB 2583 AGTCTCTTGGCTTTGTTAAAGAAAGTCAACCGGCTCCTATA 2624  
 RESULT 21  
 AAV27356  
 ID AAV27356 standard; DNA; 2290 BP.  
 XX

AAV27356;  
 02-OCT-1998 (first entry)  
 Streptococcus pneumoniae SP0042 nucleotide.  
 Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 detection; pneumonia; otitis media; meningitis; ss.  
 Streptococcus pneumoniae.  
 Key Location/Qualifiers  
 CDS 2..2290  
 /tag= a  
 /product= "SP0042"  
 /transl\_except= (pos:152..154,aa:Xaa)  
 /transl\_except= (pos:1406..1408,aa:Xaa)  
 /transl\_except= (pos:1430..1432,aa:Xaa)  
 /note= "no stop codon given; Xaa is unspecified"  
 WO9818930-A2.  
 07-MAY-1998.  
 30-OCT-1997; 97WO-US019422.  
 31-OCT-1996; 96US-0029960P.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Kunsch CA, Choi GH, Johnson LS, Hromockyj A;  
 WPI; 1998-272224/24.  
 P-PSDB; AAW55095.  
 Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
 - or their epitope-containing fragments, useful in protective or  
 therapeutic vaccines, and for diagnosis.  
 Claim 1; Page 61-62; 118pp; English.  
 The present sequence encodes a protein from Streptococcus pneumoniae. The  
 nucleic acid sequence encoding the Streptococcus pneumoniae protein can  
 be useful in vaccines for inducing protective antibodies against  
 Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 are used to detect Streptococcus infection (by usual hybridisation or  
 amplification methods), also for isolating Streptococcus genes or their  
 allelic variants. The protein can be used similarly to detect specific  
 antibodies in standard immunoassays, especially for diagnosing or  
 monitoring infections. Antibodies which bind the protein are used to  
 detect corresponding antigens, to purify the protein and for passive  
 immunisation (optionally coupled to a toxin). Vaccines are administered,  
 e.g. by injection, orally or through the skin, typically at 0.01-1000  
 (especially 10-300) mu g/ml per dose  
 Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;  
 Alignment Scores:  
 Pred. No.: 2,44e-176 Length: 2290  
 Score: 2649.50 Matches: 519  
 Percent Similarity: 76.51% Conservative: 77  
 Best Local Similarity: 66.62% Mismatches: 130  
 Query Match: 63.61% Indels: 53  
 DB: 2 Gaps: 8  
 US-09-765-271-56 (1-796) x AAV27356 (1-2290)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 DB 5 TCCTATGAACCTTGGTCGTCAACCAAGCTGGTCAGGTTAAGAAAGAGCTCAATCGAGTTTCT 64  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39



Db 65 TATATAGTGTGATCAGCTGGTCAAAAGCAGCAAAACCTTGACACCAAGATGAAGTCAGT 124  
Qy 40 LysArgGluGlyLeuAsnAlaGluGlnLeuValleLysLysLeuThrAspGlnGlyTyrVal 59  
Db 125 AAGAGGAGGGGATCAACGCCGAAACAAATNGTATCATCAAGATTAAGATCAAGGTTATGTG 184  
Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaLeuLeu 79  
Db 185 ACCCTCATGAGACCATATCATATATATATGCAAGGTTCCCTTAATGATGCCATCATC 244  
Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspLeuValAsn 99  
Db 245 AGTGAAGAGCTCTCATGAAGATCGGAATATATCATGTTGAAGGATTCAGACATTTCAAT 304  
Qy 100 GluValLysGlyGlyTyrValleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119  
Db 305 GAAATCAAGGGTGGTATGTATTAAGGTAAACCGGTAAATATATATGTTAAAGGAT 364  
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluLeuAsnArgGlnLysGlnGluHis 139  
Db 365 GCAGCTCATGGGATTAATTCGGHCAAAAGAGAGATTAAACGTCAGAACGAGAAACGC 424  
Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
Db 425 AGTCATATCATCACTCA-----AGAGCAGATATATGCTGTGTCAGCCAGAGCC 475  
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrLeuPheAsnAlaSerAspLeuLeuGluAsp 179  
Db 476 CAAGACGCTTATACACCGATGATGGGTATATCTTCAATGTCATCTGATATCATCAGGAC 535  
Qy 180 ThrGlyAspAlaTyrLeuValProHisGlyAspHisTyrHisTyrLeuProLysAsnGlu 199  
Db 536 ACGGTGARGCTTATATGCTTCTCAGCGCAGCATTAACCATTAATCTTCAAGATGAG 595  
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
Db 596 TTATCAGCTAGCAGGTAGCTGCTGCAGAGCCTATGGAATGG-----640  
Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
Db 641 -----AAGCAGGATCTCGCTCTCTCTCAAGTTCTAGTTATATATGCA 682  
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
Db 683 AATCAGCTCAACCAAGATTGTCAGAACCACTGACTGTCACCTCACTTATCAT 742  
Qy 260 GlnAlaSerGlnSerAsnAspLeuSerLeuLysGlnLeuTyrLysLeuProLeu 279  
Db 743 CAA---AATCAAGGGGAAACATTTCAAGCCTTTACGTGAATTTGATGCTAAACCCCTTA 799  
Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnLeuThrSerArg 299  
Db 800 TCAGAACGCCCATGTGAAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGA 859  
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheLeuProTyrSerGln 319  
Db 860 ACGCCAGAGGTAGCTGTCCTCATGTGTAAACATTACCATTTATCTTATCCCTTATGAACA 919  
Qy 320 MetSerGluLeuGluGluArgLeuAlaArgLeuPheLeuProLeuArgTyrArgSerAsnHis 339  
Db 920 ATGCTCTGAATGGAACAAAGATTTGCTGTATTTATTTCCCTTATTTCCCTTATGTTCAAC 979  
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
Db 980 TGGGTACCAAGATTCAGACAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039  
Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
Db 1040 AGTCCGCAACCTGCACCAATCTCAACAGCTCCCAAGCAATCCAATTCAT-----GAG 1093  
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGlyLysGly 393

1094 AAATGGTCAAAAGAGCTGTTTCGAAAAAGTAGAGCGATGGTTATGCTTTTGAGGAGAAATGGA 1153  
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
Db 1154 GTTCTCTGTTATATCCAGCCCAAGATCTTTCAGCAGAAAACAGCAGCAGGATTCATGATGC 1213  
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
Db 1214 AAATGGCCAAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAATGACCTCCCA 1273  
Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
Db 1274 TCTAGTGTAGCAGAAATTTTACAATAAGGCTTATGACTTACTAGCAGAAATTCACCAAGAT 1333  
Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg 473  
Db 1334 TTACTTGATATTAAGGTCGACAAAGTTGATTTGAGGCTTTGGATAACCTGTTGGAACGA 1393  
Qy 474 LeuAsnAspGluSerThrAsnLysGlnLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
Db 1394 CTCAAGGATGTCNCAAGTGATAAAGTCAAGTTAGTGGAGATATTTCTTGCCTTCTTAGCT 1453  
Qy 494 ProIleThrHisProGluArgGluGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
Db 1454 CGGATTCGTCTATCCAGAACGTTTAGGAAAACCAAAATGCGCAATTTACCTACTGATGAT 1513  
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
Db 1514 GAGATTCAGTAGCCAAAGTTGCGGCAAGTACACACAGAGACGGTTATATCTTTGAT 1573  
Qy 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
Db 1574 CCTCGTATATAAACCAAGTGATGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 1633  
Qy 554 HisTyrIleGlyLysAspSerLeuSerLysGluLysValAlaAlaGlnAlaTyrThr 573  
Db 1634 CACTCGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGCGGAGCCAGGCTTATGCT 1693  
Qy 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
Db 1694 AAAGAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAAATCTAGGCA 1753  
Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613  
Db 1754 AAAGAGCAGAGAGCTATCTACAAACCGCTGAAAGCAGCTAAGAGAGCTGACCTGATCGT 1813  
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleProHis 633  
Db 1814 ATGCCCTTACAATCTTCAATATATCTGTAGAAGTCAAAACCGTAGTTTATCATCATCTCAT 1873  
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
Db 1874 TATGACCATTAACATAACATCAAAATTTGAGTGGTTTGACGAAAGCGCTTTATGAGGACCT 1933  
Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
Db 1934 AAGGGGTATACTCTTGAGGATCTTTGCGCATCTGCAAGTACTACTGTGCAACATCCAAAC 1993  
Qy 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
Db 1994 GAACGTCCCATTCAGATTAATGTTTGTAAACGTAGCAGACCATGTTCAGAAACAAA 2053  
Qy 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu-----708  
Db 2054 AATGGTCAAGCTGATACCAATCAAAACGGAACCAACGAGGAGAGAAACCTCAGACAGAA 2113  
Qy 709 ---ProValGluGluThr-----713  
Db 2114 AAACCTGAGAGAAACCCCTCGAGAGAAACCCGCAAGACGAGAAACCCAGAGTCTCCA 2173  
Qy 714 -----ProAlaGluProGluValProGlnValGluThr 724  
Db 2174 AAACCAACAGAGAACCAAGAGATCAACAGAGGATCAGAAAGAACCTCAGGTCGAGACT 2233



QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743  
 DB 2234 GAAAGAGTTGAGAAAAAAGCTGAGAGAGGCTGAAGATTTCCTGGAAAAAATCCAGGAT 2290  
 RESULT 22  
 ABQ84824  
 ID ABQ84824 standard; DNA; 2290 BP.  
 XX AC ABQ84824;  
 XX DT 04-SEP-2002 (first entry)  
 XX DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.  
 XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection; gene; ds.  
 XX OS Streptococcus pneumoniae.  
 XX PN US2002061545-A1.  
 XX PD 23-MAY-2002.  
 XX PF 22-JAN-2001; 2001US-00765272.  
 XX PR 30-OCT-1997; 97US-00961083.  
 XX PA (CHOI/) CHOI G H.  
 XX PA (KUNS/) KUNSCH C A.  
 XX PA (BARA/) BARASH S C.  
 XX PA (DILL/) DILLON P J.  
 XX PA (DOUG/) DOUGHERTY B.  
 XX PA (FANN/) FANNON M R.  
 XX PA (ROSE/) ROSEN C A.  
 XX PI Choi GH, Kunsch CA, Barash SC, Dillon PU, Dougherty B, Fannon MR;  
 XX PI Rosen CA;  
 XX FI WPI: 2002-479261/51.  
 XX DR P-PSDB; ABP54589.  
 XX DR  
 XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus  
 PT and for preventing or attenuating disease caused by Streptococcus  
 PT infection.  
 XX PS Claim 1; Page 28-29; 70pp; English.  
 XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.  
 CC pneumoniae antigens have antibacterial activity and can be used in  
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or  
 CC attenuate a Streptococcal infection in an animal. The polynucleotides  
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus  
 CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning  
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example  
 CC from the present invention  
 XX SQ Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;  
 Alignment Scores:  
 Pred. No.: 2,44e-176 Length: 2290  
 Score: 2649.50 Matches: 519  
 Percent Similarity: 76.51% Conservative: 77  
 Best Local Similarity: 66.62% Mismatches: 130  
 Query Match: 53  
 DB: 6 Gaps: 8  
 US-09-765-271-56 (1-796) x ABQ84824 (1-2290)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 DB 5 TCCTATGAACCTTGGTCGTCACCAAGCTGTCAGGTTAAGAAAGAGTCTAATCGATTCT 64

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 DB 65 TATATAGTGTGATCAGCTGGTCAAAAGGCAGAAAACTTGACACACAGATGAAGTCAGT 124  
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 DB 125 AAGAGGGAGGGGATCAACGCCGAACAATNGTATCAAGATTACGGATTCAAGGTTATGTG 184  
 QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 DB 185 ACCTCTCATGGAGACCATTAATCTATATAATGCGAAGGTTCTTATCATGCCATCATC 244  
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 DB 245 AGTGAAGAGCTCTCATGAAGATCCGAATTAATCAGTTGAAGGATTACAGACATTGTCAAT 304  
 QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
 DB 305 GAAATCAAGGTGGTTATGTCAATTAAGGTAAACGGTAAATACTATGTNTACCTTAAGGAT 364  
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 DB 365 GCAGCTCATCGGATATAATTCGACAAAGAGAGATTAAACGTCAGAGCAGAACGCC 424  
 QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 DB 425 AGTCATAATCACTCACTCA-----AGAGCAGATAATGCTGTGTCAGCCAGAGCC 475  
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 DB 476 CAGGACGTTATACACGGATGATGGGTATATCTTCAATCATCATCTGATATCAATGAGGAC 535  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 DB 536 ACGGTGATGCTTATATCGTTCCTCAGCGCGACCATTAACATTACATTCTCTAAGAAATGAG 595  
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 DB 596 TTATCAGCTAGCAGATTAGCTGCTGCAGAGCCCTATTGGAATGGG----- 640  
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 DB 641 -----AAGCAGGGATCTCGCTCTTCAAGTCTTAGTTATTAATGCA 682  
 QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 DB 683 AATCCAGCTCAACCAAGATTGTCAGAGAACCACTGCTGACTGCTCACTCAACTTATCAT 742  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 DB 743 CAA---AATCAAGGGGAAACAACTTCAAGCCTTTTACGTGAATTTGTATGCTAAACCCCTTA 799  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 DB 800 TCAGAACGCATGTGGAATCTGATGGCCCTTATTTTCGACCAGCGCAAAATCACAAGTCGA 859  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 DB 860 ACCGCCAGAGGTAGTGTGCTCCCTCATGTGTAACCATTAACCATTTATCTCTTATGACAA 919  
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 DB 920 ATGTCCTGAATTGGAAAAAGCAATTCGTATTAATTCCTCCCTTCCTTATCGTTCAAAACCAT 979  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 DB 980 TGGGTACCAAGATTCAAGACCAAGAACCAAGTCCCAATTCGACTCCGGAACCTAGTCCA 1039  
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 DB 1040 AGTCCGCAACCTGCACCAATCTCTCAACGAGCTCCAGCAATCCCAATTGAT-----GAG 1093

QY	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393	Db	2174	AAACCAACAGAGGAACCCAGAGAATCCACGAGGAATCAGAGAACTCCAGTTCGAGACT	2233
Db	1094	AAATGGTCAAGAGAGCTGTCGAAAGTAGAGCGATGTTGTTATGCTTTGAGAGAGATGGA	1153				
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413	QY	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAsp	743
Db	1154	GTTCTCTGTTATATCCAGCCCAAGGATCTTTCAGCAGAGAAACAGCAGCAGGCGATTCATAGC	1213	Db	2234	GAAAAGGTGAAGAACTGAGAGAGCTGAAAGATTTTCTTGGAAAAATCCAGGAT	2290
QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433	RESULT 23			
Db	1214	AAACTGGCCAAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAACTGACCTCCA	1273	ADC45146			
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453	ID	ADC45146	standard; DNA; 2290 BP.	
Db	1274	TCTAGTATCGAAGATTTTACATTAAGCTTATGAGCTTCTAGCAAGATTCACCAAGAT	1333	XX	ADC45146;		
QY	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg	473	XX	18-DEC-2003	(first entry)	
Db	1334	TTACTTGTATATAAAGGTCGACAACTTCATTTGAGGCTTTGGATAAAGCTGTTGGAACGA	1393	XX	S. pneumoniae	DNA encoding antigen SP042.	
QY	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAla	493	XX	Antigen; ds; bacterial infection; vaccine; pneumonia; antibacterial.		
Db	1394	CTCAAGGATGTCNCAAGTGATAAAGTCAGGATGAGTGTAGTGAGATATTTTGCCTTCTTAGCT	1453	XX	Streptococcus pneumoniae.		
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513	XX	US6573082-B1.		
Db	1454	CGATTCGTATCCAGAACGTTTAGAAAAACCAATGCGCAATTTACCTACATCATGAT	1513	XX	03-JUN-2003.		
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAsp	533	XX	28-MAR-2000; 2000US-00536784.		
Db	1514	GAGATTCAAGTAGCCCAAGTTGCGCAGCAAGTACACACAGAGAAGCGTTATATCTTTGAT	1573	XX	31-OCT-1996; 96US-0029960P.		
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553	XX	30-OCT-1997; 97US-00961083.		
Db	1574	CCTCGTGATATACACAGTGATGAGGGGAGTCTATGTAACTCCACATATGACCCATAGC	1633	XX	(HUMA-) HUMAN GENOME SCI INC.		
QY	554	HisTyrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573	XX	Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;		
Db	1634	CACTGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGCGCGACGCCAGCTTATGCT	1693	XX	Rosen CA;		
QY	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593	XX	WPI; 2003-764574/72.		
Db	1694	AAAGAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGGAATACTCAGGCA	1753	XX	P-PSDB; ADC45147.		
QY	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg	613	XX	Novel polynucleotide encoding Streptococcus pneumoniae polypeptides		
Db	1754	AAAGGAGCAGAGCTATCTACAAACCGCGTGAAGCAGCTAAGAAAGTGCCATTCATGCT	1813	XX	useful for producing vaccines for prevention or attenuation of infection		
QY	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633	XX	Example 1; SEQ ID NO 65; 58pp; English.		
Db	1814	ATGCCCTTACAATCTTCAATATCTAGTAAGTCAAAACCGTAGTTTATCATCATCTCAT	1873	XX	The invention relates to an isolated polynucleotide consisting of a		
QY	634	LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro	653	XX	Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding		
Db	1874	TATGACCATTACCATCAATCAAAATTTGAGTGGTTTGAACGAGCCCTTTATGAGGACCT	1933	XX	SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae		
QY	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrThrValGluHisProAsp	673	XX	nucleic acid into a vector, an isolated polynucleotide consisting of at		
Db	1934	AAGGGTATATCTTCTGAGGATCTTTGGCGACTGTCAAGTACTAAGTCAATCCAACTCCAAAC	1993	XX	least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a		
QY	674	GluArgProHisSerAsnAspGlyTyrPheAsnAlaSerGluHisValLeuGlyLysLys	693	XX	recombinant host cell comprising the SP028 polynucleotide. The nucleic		
Db	1994	GAAGCTCCGCANTTCAGATAATGTTTTTGTAAACGCTAGCAGCACCATTCTCAAGAAACAAA	2053	XX	acids are useful as DNA vaccine against Streptococcus pneumoniae		
QY	694	AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu	708	XX	infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae		
Db	2054	AATGGTCAAGCTGATACCAATCAACCGAAAAACCAAGCGAGGAGAAACCTCAGACAGAA	2113	XX	antigen nucleic acids are useful as probes for use in diagnostic methods		
QY	709	---ProValGluGluThr	713	XX	for detecting S. pneumoniae gene expression. The present sequence encodes		
Db	2114	AAACCTGAGAGAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCA	2173	XX	an S. pneumoniae antigenic protein.		
QY	714	-----ProAlaGluProGluValProGluValGluThr	724	XX	Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 0 U; 5 Other;		
				Alignment Scores:			
				Pred. No.:	2,44e-176	Length:	2290
				Score:	2649.50	Matches:	519
				Percent Similarity:	76.51%	Conservative:	77
				Best Local Similarity:	66.62%	Mismatches:	130
				Query Match:	63.61%	Indels:	53
				DB:	9	Gaps:	8
				US-09-765-271-56 (1-796) x ADC45146 (1-2290)			
QY				1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer	19	
Db				5 TCTATGAATCTTGCTGCTCACCAGCTGCTCAAGAGAGTCTAATCGAGTTCT	64		

QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
DB 65 TATATAGATGGTATCAGGCTGGTCAAAAGGCAGAAAACCTTGACACAGATGAAGTCAGT 124  
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
DB 125 AAGAGGAGGGGATCAACGGCGAAACAAATNGTATCAAGATTACGGATCAAGGTTATGTG 184  
QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
DB 185 ACCTCTCATGGAGACCAATATCATTTACTATATGCGCAAGGTTCCTTATGATGCCATCATC 244  
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
DB 245 AGTCAAGAGCTCTCATGAAGATCCGAATTCAGTTGAAGATTTCAGACATTTGTCAAT 304  
QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
DB 305 GAAATCAAGGGTGGTTATGTTCATTAAGGTAAACGGTAAATACTATGTTNACCTTAAGGAT 364  
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
DB 365 GCAGCTCATGCGGATAATATTCGACAAAAGAGAGATTAAACGTCAGAAGCAGAAACGC 424  
QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
DB 425 AGTCATAATCATTAATCA-----AGACGAGATAATGCTGTGTCGAGCCAGAGCC 475  
QY 160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
DB 476 CAAGGAGGTTTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGCAC 535  
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
DB 536 ACGGGTATGCTTATATCGTTCCTCAGCGGACCAATACCATTAACATTCCTTAAGAATGAG 595  
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
DB 596 TTATCAGTAGCAGTTAGTGTGTCGAGAGCCTATTGGATGG----- 640  
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
DB 641 -----AAGCAGGATCTCGTCTCTTCAAGTCTAGTTATTAATGCA 682  
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
DB 683 AATCCAGCTCAACAGATTGTGAGAACCAACAAATCTGACTGCTCAACCTTATCAT 742  
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
DB 743 CAA--AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCCCTTA 799  
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
DB 800 TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAACAAGTCGA 859  
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
DB 860 ACCGCAGAGGTGAGTGTCCCTCATGGTAACCAATACCACTTTATCCCTTATGAACAA 919  
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
DB 920 ATGCTGAATTGGAAAAACGAATTTGCTGTATTTATTTCCCTTCTGTTATCTTCAACCAT 979  
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
DB 980 TGGGTACCAATCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039  
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
DB 1040 AGTCCGCAACCTGCACCAATTCCTCAACCAAGCTCCAGCAATCAATTTGAT-----GAG 1093  
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393

DB 1094 AATTTGTCAAAGAGCTGTTTCGAAAAGTAGCGGATGTTTGTATGCTTTGAGGAGAAATGA 1153  
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
DB 1154 GTTTCCTGTTATATCCAGCCCAAGGATCTTTCAGCAGAAAACAGCAGCAGGCTATTGATAGC 1213  
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
DB 1214 AAATCGCCCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAAACCTGACCTCCCA 1273  
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
DB 1274 TCTAGTGATCGAGAAATTTTCAATAGCTTATGACTTACTAGCAAGAAATTCACCAAGAT 1333  
QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
DB 1334 TTACTTGATAATAAAGGTCGACCAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGA 1393  
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
DB 1394 CTCAAGGATGTCNCAAGTGATAAAGTCAAGTTAGTGGANGATATTTCTTGCTCTTCTAGCT 1453  
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
DB 1454 CCGATTCTGTCATCCAGAACGTTTAGGAAAACCAATGCGCAATATTACTACACTGATGAT 1513  
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
DB 1514 GAGATTCAAGTAGCAAGTTGGCAGGCAAGTACACACAGAACGCGTTATATCTTTGAT 1573  
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
DB 1574 CCTCGTGATATAACCAAGTGATGAGGGGATGCTTATGTAATCCACATATAGCCCATAGC 1633  
QY 554 HistripleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
DB 1634 CACTGGATTAANAAGATAGTTTGTCTGAAGCTGAGAGAGCGGAGCCAGGCTTATGCT 1693  
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
DB 1694 AAAGAGAAAGTTTGAACCTCTCTCGACAGACCATCAGGATTTCAGGAAATACTGAGGCA 1753  
QY 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613  
DB 1754 AAAGGAGCAGAAAGCTATCTACAACCGCGTGAAGCAGCTAAGAAGGTGCCACTGATCGT 1813  
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633  
DB 1814 ATGCTTACATCTTCAATATCTGTAGAAGCTCAAAAACGCTAGTTTATCATCTACCTCAT 1873  
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
DB 1874 TATGACCAATTACCATTAACATCAAAATTTGAGTGGTTTGACGAAGCGCTTTATGAGGCACCT 1933  
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
DB 1934 AAGGGGTATCTCTTGAAGGATCTTTTGGCGACTGTCAAGTACTATGTGAAACATCCAAAC 1993  
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
DB 1994 GAACGTCGCGATTCAAGATTAATGGTTTTTGGTAACGCTAGCAGCAATGTTTCAAGAAACAA 2053  
QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708  
DB 2054 AATGGTCAAGCTCATACCAATCAACCGAAAAACCAAGCGGAGGAGAAACCTCAGACAGAA 2113  
QY 709 ---ProValGluGluThr----- 713  
DB 2114 AAACCTGAGGAAGAAACCCCTCGAGAAGAGAAACCGCAAGCGAGAGAACCCAGAGTCTCCA 2173  
QY 714 -----ProAlaGluProGluValProGlnValGluThr 724



```
Db 1033 TTTTACAATAAGGCTTATGACTTACCAAGAAATTCACCAAGATTTTACTTGATAATAAA 1092
Qy GlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSer 478
Db 1093 GGTGCAACAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGACGACTCAAGGATGTCCTCA 1152
Qy ThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisPro 498
Db 1153 AGTGATAAAGTCAAGTTAGTGATGATATCTTCCCTTCTTAGCTCCGATTCGTCATCCA 1212
Qy GluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAla 518
Db 1213 GAACGTTTAGGAAACCAAAATGCGCAAATTAACCTACACTGATGATGAGATTCAAGTAGGCC 1272
Qy GlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIle 538
Db 1273 AAGTTGCGCAGGCAAGTACACAAACAGAGACGGTTATATCTTTGATCCTCGTGATATAACC 1332
Qy SerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLys 558
Db 1333 AGTGATGAGGGGATGCTATGTAACCTCCACATATGACCCATGACTGGATTAAAAAA 1392
Qy AspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIle 578
Db 1393 GATAGTTTGTGAAGCTGAGAGCGCAGCCAGGCTTATGCTAAAGAGAAAGGTTTG 1452
Qy LeuProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAla 598
Db 1453 ACCCTCTCTCGACAGACCAATCAGGATTCAGGAATAATCTGAGGCAAAAGGAGCAGAGCT 1512
Qy IleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetVal 618
Db 1513 ATCTACACCGGTGAAGCAGCTAAGAGGTGCATTTGATCGTATGCTTACATCTT 1572
Qy GluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHis 638
Db 1573 CAATATACTGTAGAAGTCAAAACCGTAGTTTAAATCATCTATTATGACCATTACCAT 1632
Qy AsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeu 658
Db 1633 AACATCAAAATTTGAGTGGTTTGACGAAGGCGCTTTATGAGGCACCTAAAGGGGTATACCT 1692
Qy GluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSer 678
Db 1693 GAGATCTTTGGCATCTGCAAGTACTATGTGCAATCATCAACGAAACGTCGCGCATTC 1752
Qy AsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAsp 698
Db 1753 GATAATGGTTTGTGAACGCTAGCGACCATGTTTCAAAAGAAACAAATGTCAGCTGAT 1812
Qy ProAsnLysAsnPheLysAlaAspGluGlu-----ProValGluGlu 712
Db 1813 ACCAATCAAAACGGAACCAACGAGGAGAAACCTCAGACAGAAACCTGAGGAGAA 1872
Qy Thr----- 713
Db 1873 ACCCTCGAAGAGAAACCGCAAGCGAGAAACCAAGAGTCTCCAAACCAACAGAGGAA 1932
Qy -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAla 729
Db 1933 CCAGAGAATCACCAGAGGAATCAGAGAACCTCAGGTCGAGACTGAAAGGTTGAAGAA 1992
Qy GlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsn 749
Db 1993 AAATCTGAGAGGCTGAAGATTTTACTTGGAAATAATCCAGGATCCAAATTAATCAAGTCCAAT 2052
Qy AlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsn 769
Db 2053 GCCAAGAGACTCTCAGAGATTTAAAAAATAATTTACTATTGCGCACCAGGACCAACAT 2112
Qy SerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
:::|||||
```

```
Db 2113 ACTATTATGCGAGAAGCTGAAAACTATTGCTTTATTAAAGGAGAGT 2160
RESULT 25
AAV52376
ID AAV52376 standard; DNA; 2359 BP.
XX
AC AAV52376;
XX
XX 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:243.
XX
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX computer readable medium; vaccine; pharmaceutical composition; ds.
XX
XX Streptococcus pneumoniae.
XX
XX WO9818931-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019588.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
XX Dougherty BA;
XX
XX WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
XX Claim 1; Page 1265-1266; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has the
XX nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
XX 391, identifying members of the library which contain sequences that
XX hybridize to the target sequence and isolating the nucleic acid molecules
XX from the members; or (b) isolating mRNA, DNA or cDNA produced from an
XX organism, amplifying nucleic acid molecules whose nucleotide sequence is
XX homologous to amplification primers derived from the fragment of the S.
XX pneumoniae genome to prime the amplification and isolating the amplified
XX sequences. The computer readable medium can be used in a computer-based
XX system for identifying fragments of the S. pneumoniae genome of
XX commercial importance, or expression modulating fragments of the S.
XX pneumoniae genome. Products from the present invention can be used in
XX diagnosis kits and assays, and pharmaceutical compositions and vaccines
XX for S. pneumoniae
XX
XX Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 6,07e-115 Length: 2359
Score: 1774.00 Matches: 347
Percent Similarity: 81.22% Conservative: 51
Best Local Similarity: 70.82% Mismatches: 74
Query Match: 42.59% Indels: 18
DB: 2 Gaps: 5
```

US-09-765-271-56 (1-796) x AAV52376 (1-2359)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 1, 2004, 07:38:42 ; Search time 4946 seconds  
(without alignments)  
4805.964 Million cell updates/sec

Title: US-09-765-271-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKENRVSY.....KLLALLKGSNPSSVSKKIN 796

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US09765271/runat\_30092004\_113754\_12290/app\_query.fasta\_1.967  
-DB-EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=90  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09765271@cgn\_1\_1\_3609@runat\_30092004\_113754\_12290 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estom:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vri:\*  
28: gb\_gsal:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177	4.2	180	28	AF325819	AF325819 AF325819
2	155	3.7	2232	11	BC016682	BC016682 Homo sapi
3	144	3.5	2754	29	AY416216	AY416216 Mus muscu
4	144	3.5	3165	11	BC029682	BC029682 Mus muscu
5	144	3.5	3166	11	AK088741	AK088741 Mus muscu
6	144	3.5	3706	11	BC049893	BC049893 Mus muscu
7	142	3.4	2581	11	AK003180	AK003180 Mus muscu
8	141	3.4	2532	11	AK085512	AK085512 Mus muscu
9	140.5	3.4	2740	11	AK029960	AK029960 Mus muscu
10	139.5	3.3	3045	11	BC014691	BC014691 Mus muscu
11	139.5	3.3	4984	11	AY325172	AY325172 Rattus no
12	139	3.3	3573	29	AY404467	AY404467 Pan trogl
13	138.5	3.3	1831	29	CG756440	CG756440 P051-3-G1
14	138	3.3	3313	11	AK029864	AK029864 Mus muscu
15	137	3.3	2582	11	AK016539	AK016539 Mus muscu
16	137	3.3	2757	29	AY416214	AY416214 Homo sapi
17	137	3.3	5304	11	BC048778	BC048778 Mus muscu
18	135	3.2	2136	29	CG756344	CG756344 P051-3-E1
19	135	3.2	4561	11	AK054014	AK054014 Mus muscu
20	134.5	3.2	1643	29	CG753778	CG753778 P049-1-A0
21	134.5	3.2	1652	29	CG756662	CG756662 P051-4-D0
22	134.5	3.2	1755	29	CG754413	CG754413 P049-4-D0
23	134.5	3.2	4773	29	AY411030	AY411030 Homo sapi
24	134	3.2	2041	29	CG755974	CG755974 P051-2-E1
25	134	3.2	3714	29	AY404466	AY404466 Homo sapi
26	132.5	3.2	1948	11	AK014892	AK014892 Mus muscu
27	132.5	3.2	2791	11	AK081763	AK081763 Mus muscu
28	132	3.2	3526	11	AK054172	AK054172 Mus muscu
29	132	3.2	6914	29	AY418553	AY418553 Mus muscu
30	131.5	3.2	1787	29	CG754239	CG754239 P049-3-D1
31	130	3.1	3591	29	AY401924	AY401924 Mus muscu
32	129.5	3.1	1439	11	AK077177	AK077177 Mus muscu
33	129.5	3.1	1647	29	CG756464	CG756464 P051-3-H0
34	129.5	3.1	2950	11	AK054096	AK054096 Mus muscu
35	129	3.1	1744	29	CG755895	CG755895 P051-2-D0
36	129	3.1	2934	11	AK089964	AK089964 Mus muscu
37	129	3.1	4857	11	AK045314	AK045314 Mus muscu
38	128.5	3.1	1966	29	CG756138	CG756138 P051-3-A0
39	128.5	3.1	4358	11	BC035041	BC035041 Homo sapi
40	128	3.1	2294	11	AY321329	AY321329 Rattus no
41	128	3.1	2299	11	AY310142	AY310142 Rattus no
42	128	3.1	4107	29	AY420059	AY420059 Mus muscu
43	127	3.0	2067	29	CG756393	CG756393 P051-3-F1
44	126.5	3.0	1797	29	CG756576	CG756576 P051-4-B0
45	126.5	3.0	1877	11	AK077954	AK077954 Mus muscu
46	126.5	3.0	1882	29	CG755934	CG755934 P051-2-E0
47	126.5	3.0	5400	29	AY418797	AY418797 Homo sapi
48	125.5	3.0	1072	29	CNS0174M	AL166063 Tetraodon
49	125	3.0	2175	11	BC016130	BC016130 Mus muscu
50	124.5	3.0	2748	11	AK082361	AK082361 Mus muscu
51	124.5	3.0	5438	29	AY402980	AY402980 Mus muscu
52	124.5	3.0	6836	29	AY405716	AY405716 Mus muscu
53	124.5	3.0	9744	29	AY405009	AY405009 Homo sapi
54	124	3.0	954	11	CNS08XNP	BC033041 Single re
55	124	3.0	1783	29	CG754377	CG754377 P049-4-B0
56	123.5	3.0	1019	28	AZ685876	AZ685876 ENTH1177F
57	123.5	3.0	1962	29	CG756438	CG756438 P051-3-G1
58	123.5	3.0	2595	11	AK028829	AK028829 Mus muscu
59	123.5	3.0	2884	11	AK029095	AK029095 Mus muscu
60	123.5	3.0	2981	11	AK004763	AK004763 Mus muscu
61	123.5	3.0	2994	11	AK030223	AK030223 Mus muscu
62	123	3.0	3582	11	AK031955	AK031955 Mus muscu



```

63      123      3.0      3991      11      AK036731      Mus muscu
64      122.5      2.9      1071      10      BF237738      601842247
65      122.5      2.9      1505      20      BG024255      602302990
66      122.5      2.9      2577      29      AY406623      Homo sapi
67      122.5      2.9      3209      11      AK046752      Mus muscu
68      122.5      2.9      3541      11      AK030766      Mus muscu
69      122.5      2.9      4537      29      AY417627      Homo sapi
70      122      2.9      1785      29      CG750213      P044-3-H0
71      122      2.9      3198      11      BC009572      Homo sapi
72      121.5      2.9      1767      29      CG756340      P051-3-E0
73      121.5      2.9      2038      11      AK089164      Mus muscu
74      121.5      2.9      2218      12      BG169169      602320648
75      121.5      2.9      2380      11      AK013754      Mus muscu
76      121.5      2.9      2571      11      AK028247      Mus muscu
77      121.5      2.9      3110      11      AK019964      Mus muscu
78      120.5      2.9      1812      12      BI655534      603285602
79      120.5      2.9      1896      29      CG755802      P051-2-B0
80      120.5      2.9      3626      11      AK054303      Mus muscu
81      120      2.9      921      9      AL519296      AL519296
82      120      2.9      1515      10      BF206929      601870371
83      120      2.9      2413      11      AK019969      Mus muscu
84      120      2.9      2893      29      AY401552      Mus muscu
85      120      2.9      4137      29      AY420057      Homo sapi
86      119.5      2.9      714      12      BJ350580      BJ350580
87      119.5      2.9      1173      29      CNS06TPT      T3 end of
88      119.5      2.9      1599      10      BF981422      602309223
89      119.5      2.9      1733      29      CG756798      P051-4-G0
90      119.5      2.9      1823      29      CG756266      P051-3-D0

```

# ALIGNMENTS

```

RESULT 1
AF325819      AF325819      180 bp      DNA      linear      GSS 08-MAR-2002
LOCUS      AF325819      Leptospira interrogans serovar lai DNA Leptospira
DEFINITION      Leptospira interrogans serovar lai genomic clone 18, genomic survey sequence.
ACCESSION      AF325819
VERSION      AF325819.1      GI:19263459
KEYWORDS      GSS.
SOURCE      Leptospira interrogans serovar lai
ORGANISM      Leptospira interrogans serovar lai
REFERENCE      1 (bases 1 to 180)
AUTHORS      Hu, C. and Bao, J.
TITLE      Leptospira interrogans serovar lai DNA
JOURNAL      Unpublished (2002)
COMMENT      Contact: Hu C
Leptospirosis Research Unit
West China University of Medical Sciences
17# Reiming Nan Road, Chengdu, Sichuan 610041, P.R. China
Email: huchanghua@263.net
Class: unknown.

```

```

FEATURES
    source
        1..180
            /organism="Leptospira interrogans serovar lai"
            /mol_type="genomic DNA"
            /db_xref="taxon:57678"
            /clone="18"
            /clone_lib="Leptospira interrogans serovar lai DNA"
            /note="Obtained through DNA subtraction using the
nonpathogenic strain Leptospira biflexa serovar Patoc
strain Patoc I"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      4.27e-05      Length:      180
Score:      177.00      Matches:      31
Percent Similarity:      74.55%      Conservative:      10
Best Local Similarity:      56.36%      Mismatches:      14
Query Match:      4.25%      Indels:      0
DB:      28      Gaps:      0

```

```

US-09-765-271-56 (1-796) x AF325819 (1-180)
Qy      84      LeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
Db      3      TTAATGACGACCCCAATTATCAGTTTAAACAGTCTGATGTTGTTTAAACGAGATTCTAGAT 62
Qy      104      GlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAlaAlaHisAla 123
Db      63      GGCTATATATCAAGTTGATGCAAGATTATTTATTTATTAAGCCAGGAAGCAAGCGT 122
Qy      124      AspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGlu 138
Db      123      AAAAATATCCGTACCAACAGCAAAATTCCTGAACAGGTGAAAAG 167

```

## RESULT 2

```

BC016682      Homo sapiens, Similar to KIAA0164 gene product, clone
LOCUS      IMAGE:3923680, mRNA.
DEFINITION      BC016682.1      GI:16741778
ACCESSION      BC016682
VERSION      BC016682
KEYWORDS      HTC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 2232)
AUTHORS      Strausberg, R.
TITLE      Direct Submission
JOURNAL
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/PTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalalon@bcm.tmc.edu.
Villalalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX plate: 26 Row: d Column: 7  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7661957  
This clone has the following problem: frame shifted.

## FEATURES

```

    Location/Qualifiers
        1..2232
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3923680"
            /tissue_type="Skin, melanotic melanoma."
            /clone_lib="NIH MGC_72"
            /lab_host="DH10B"
            /note="Vector: pCMV-SPORT6"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      0.12      Length:      2232
Score:      155.00      Matches:      137
Percent Similarity:      32.67%      Conservative:      92
Best Local Similarity:      19.54%      Mismatches:      248
Query Match:      3.72%      Indels:      224
DB:      11      Gaps:      29

```



```

RESULT 3
LOCUS AY416216 2754 bp DNA linear GSS 12-DEC-2003
DEFINITION Mus musculus HCM5816 gene, VIRTUAL TRANSCRIPT, partial sequence,
GENOMIC survey sequence.
ACCESSION AY416216
VERSION AY416216.1 GI:39772176
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
location/Qualifiers
1..2754
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2754
/locus_tag="HCM5816"
gene
ORIGIN
Alignment Scores:
Pred. No.: 1.31 Length: 2754
Score: 144.00 Matches: 137
Percent Similarity: 32.7% Conservative: 94
Best Local Similarity: 19.43% Mismatches: 240
Query Match: 3.46% Indels: 234
DB: 29 Gaps: 30
US-09-765-271-56 (1-796) x AY416216 (1-2754)
QY 127 ArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGlyGly 146
Db 46 AGATCACAGCTCAGTCTTAGATCAAGATCAAGATCACATTCCTAGAAAGAGAGATACAGT 105
QY 147 ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAsp 166
Db 106 TCTAGGTCCTCGTCCAGAACATATCGAGGTCCTCGTAGTAGAGATCTATTATTCTAGA 165
QY 167 Asp-----GlyTyrIlePhe 171
Db 166 GATTATCGTCGAGATTACAGGAATAATACAGGAATGAGACGACCTTATGGGTAC----- 219
QY 172 AsnAlaSerAspIleileGluAspThrGlyAspAlaTyrIleValProHisGlyAspHis 191
Db 220 -----AGAGAGAGGGGTAGAGGGTATTATCAAGAGGAGAGGGGAGA 261
QY 192 TyrHis-----TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAla 208
Db 262 TACCATCGAGGTGGCTATAGACCTGCTCGAATAGAGGCACTCTAGG----- 309
QY 209 GluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArgGlnAsn 228

```

```

Db 310 -----AGTCCTAGACGAGGTCTCGTCCAGGTCCAGGAGTCCAAAAGAGATGCC 357
QY 229 ---SerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGlyThrThr 247
Db 358 GTGCTCTTCTCAAGATCCGAGGAGAGATCTCGCGGTCTATAGATCTCTTAGGTCTCCA 417
QY 248 AsnThrAsnThrAsnSerAsnSerGlnAlaSerGlnAlaSerGlnSerAsnAspIle 267
Db 418 AGATCATCATCATCTCGCTCTTCATCCCATATAGCAAAATCTCTGCTCTCTTAAAGACGA 477
QY 268 AspSerLeuLeuLysGlnLeuTyrIleLysLeu-----ProLeu 279
Db 478 GGGTCTCAGGAAAAACAAACCAAGCTGAGGGGGAACCCCAAGAGAGAGATCCTTTG 537
QY 280 SerGlnArgHisValGlu-----SerAspGlyLeuValPheAspProAla----- 294
Db 538 AAAAGCAAAATCAGGAGGAAACCAAGATACTTTTGAACATGATCCATCTGAATCTATT 597
QY 295 ---GlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHis 313
Db 598 GATGAGTTTAAATAAATCTGCTACTTCTGCTGATATTGCGCTGCGCTTTCAGCTTATGAT 657
QY 314 PheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleIleProLeu 333
Db 658 AATAGTCCCAAGGTCA----- 672
QY 334 ArgTyrArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGlnPro 353
Db 673 -----CCTCATAGTCTCTTCCACTTATGCT 696
QY 354 Thr---ProGluProSerProGlyProGlnProAlaProAsnLeuLysIle----- 369
Db 697 ACACCAACCCAGTCAGAGTTTCATGCTCAGATGCCCATCTGCTTAGTACAGTGCCTCT 756
QY 370 -----AspSerAsnSerSerLeuValSerGlnLeuValValArgLysValGly 384
Db 757 GCCAAAAATACCCCTTCTCAGCATTCATCCATTCAGACAGAGTCTCGAGAGGTCTGA 816
QY 385 GluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuPro 404
Db 817 TCTGCTCTGTT-----GGAAATGGGTGCGAGTCGATATAGTCTCTCAGATAGTCG 870
QY 405 SerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisThr 424
Db 871 ATTCATCATCATCCCATCA-----CGAAGAAGCCCTCCCAAGACA 909
QY 425 LeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyr 444
Db 910 ATCACACCAACAGAAAT-----GCTCCACAGAGAGAA----- 939
QY 445 AsnLeuLeuThrGluAlaHisLysAlaLeuPhe**AsnLysGlyArgAsnSerAspPhe 464
Db 940 -----TCTAGGGGGGGTCCCTCTTTAT 963
QY 465 GlnAlaLeuAsp-----LysLeuLeuGluArgLeuAsnAspGlu 477
Db 964 CCGTAAGGAGATCAGAAACCAAGACAGGAAAGTTTAAAAAGGTTCACAGATGAA 1023
QY 478 SerThr-----AsnLysGluLysLeuValAsp 486
Db 1024 GAGTCTAGAGTATTCTGCTTGTATAGGGGTAAATATCAGGGGATAAAGAG----- 1071
QY 487 AspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSer 506
Db 1072 -----GCTCCGAAGGAGAAAGGTGTCAGAAAGGCGAGGCGAGATGGA 1113
QY 507 Gln-----IleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1114 GATTGGGATGACCAAGAAATCTAGATTACTTTAGTGAATAAGAGTCTCCAAACAA----- 1170
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAsp----- 536
Db 1171 -----AAATTCATGACTCTGAAGGGGTGACACAGAGAGACAGAGGATTATAGACAG 1224

```

QY 537 -----llelleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHis 552  
 Db 1225 TTTAGGAATCAGTCTAGCAGATCAGGGCAAGAGCTTTGCTACT----- 1269  
 QY 553 SerHisTrpIleGlyLysAspSerLysSerAspLysGluLysValAlaAlaGluAlaTyr 572  
 Db 1270 -----TCATCTCACGGGAT 1284  
 QY 573 ThrLysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThr 592  
 Db 1285 ACTGAGGAGGAAGCACCACAGTACAGTCC---AAAGTTTCATTAAAGGCAATAGAGAA 1341  
 QY 593 GlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuVal 612  
 Db 1342 AGTGATGA-----TTTACAGAGAGAAAAAATTTATAAATGAAA 1380  
 QY 613 ArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlePro 632  
 Db 1381 GAGACTGCATACATAGTCCGAAGCCCTAGCACTGCAAGGAC-----AAG 1425  
 QY 633 HisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAla 652  
 Db 1426 CACAAGGAA----- 1434  
 QY 653 ProAsnGlyTyrThrLeuGluAspLeuPheAla-----ThrIleLysTyrTyrVal 669  
 Db 1435 -----GAAGCAAGGCTCTGATAGATAACAGCTGAAGAAAGAGTG 1476  
 QY 670 GluHisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisVal 689  
 Db 1477 CAGTCACCGGAGCAG-----GTAAAGTCTGAAAGCTC 1509  
 QY 690 LeuGlyLysLysAspHisSerGluAspProAsnLysAsnPhelLysAlaAspGluPro 709  
 Db 1510 AAAGAGCTCTTTGATTACAGTCCCTCTACACAGAGTCTGGACGACGAGAAAGTCC 1569  
 QY 710 Val-----GluGluThrProAlaGluProGluValProGluValGluThrGluLysVal 727  
 Db 1570 ATCTTCAGAGGAGAGCCCACTGAGATCAAAATGATAGCCAGTGTCTCACCGTCT 1629  
 QY 728 GluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLys 747  
 Db 1630 GAAGTCAAACTCAAAATGGCACCTGTTCCTTT-----GAGATTCCACAGACCT 1680  
 QY 748 AlaAsnAlaThrGlu 752  
 Db 1681 GCCTCCTTGACTAA 1695

RESULT 4  
 LOCUS BC029682 3165 bp mRNA linear HTC 07-AUG-2002  
 DEFINITION Mus musculus, similar to Bcl-2-associated transcription factor,  
 clone IMAGE:5362245, mRNA.  
 ACCESSION BC029682  
 VERSION BC029682.1 GI:20987570  
 KEYWORDS HTC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 3165)  
 Strauberg, R.  
 Direct Submission  
 Submitted (06-MAY-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-x@mail.nih.gov](mailto:cgabs-x@mail.nih.gov)  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 53 Row: D Column: 23  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Genomescan gene prediction  
 This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
 1..3165  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5362245"  
 /tissue\_type="Eye, retina, mouse strain C57Bl/6"  
 /clone\_lib="NIH\_MGC\_94"  
 /lab\_host="DH10B"  
 /notes="vector: pCMV-SPORT6"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.61 Length: 3165  
 Score: 144.00 Matches: 137  
 Percent Similarity: 32.77% Conservative: 94  
 Best Local Similarity: 19.43% Mismatches: 240  
 Query Match: 3.46% Indels: 234  
 DB: 11 Gaps: 30  
 US-09-765-271-56 (1-796) x BC029682 (1-3165)  
 QY 127 ArgThrLysGluGluLeuLeuAsnArgGlnLysGlnGluHisSerGlnHisArgGluGly 146  
 Db 227 AGATCACAGCTAGTTCTAGATCAAGATCAAGATCAATTCAGAAAGAGAGATACAGT 286  
 QY 147 ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrAsp 166  
 Db 287 TCTAGGTCTCGTTCCAGAACATATTCGAGTCTCGTAGTAGATCGTATTTTCTAGA 346  
 QY 167 Asp-----GlyTyrIlePhe 171  
 Db 347 GATTATCGTCGATATACAGGAATATAGAGGAATAGAGACCTTATGGTAC----- 400  
 QY 172 AsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHis 191  
 Db 401 -----AGAGGAAGGGGTAGAGGGTATTATCAAGGAGGAGGAGGAGA 442  
 QY 192 TyrHis-----TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAla 208  
 Db 443 TACCATCGAGGTGGCTATAGACCTGTCTGGAATAGAGGCACTCTAGG----- 490  
 QY 209 GluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArgGlnAsn 228  
 Db 491 -----AGTCTAGACGAGGTGGTCCGTCACGGTCCAGGAGTCCAAAAGAGATCC 538  
 QY 229 ---SerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGlyThrThr 247  
 Db 539 GTGTCTTCTCAAAGATCCCGAAGCAGATCTCGCGGTGTCATATAGATCTCTAGTCTCCA 598  
 QY 248 AsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIle 267  
 Db 599 AGATCATCATCATCTCGCTCTTCATCCCATATAGCAAAATCTCTCTCTCTCTCTCT 658  
 QY 268 AspSerLeuLeuLysGlnLeuTyrLysLeu-----ProLeu 279  
 Db 659 GGGTCTCAGGAACAAACCAACCAAAAGAGGTGAGGGGGAACCCCAAGAGAGATCTCTTTG 718



Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3166)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishio, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muranatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics  
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
Trust/MRC building Addenbrookes Hospital Cambridge) whose  
assistance we gratefully acknowledge.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. .3166  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="FANTOM,DB:E430024P04"  
/db\_xref="MGI:2427638"  
/db\_xref="taxon:10090"  
/clone="B430024P04"  
/cell\_type="thymic cells"  
/tissue\_type="thymus"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="2 days neonate"  
192. .2438  
/notes="unnamed protein product; hypothetical protein  
(SPTR) Q14673, evidence: rSCDS, NCBI CDS Predictor)  
putative"  
/codon\_start=1  
/protein\_id="BAC40542.1"  
/db\_xref="GI:26353824"  
/translation="MGRSRSRSHSRSSKRSOSSPSRSHSRKRYRSRTYRSR  
RSRDYRDRYDRYNNRMRPGYRGGRGYQGGGRHRCGRYRVRWRRSRS  
PRGRSRSPKRSVSQSRSRSRSRSSPSRSPSSSPSPKRSRGSQ  
EKQTKAEQPEESLKSQEPKDFEHPDSIEDFNKSSATSGDIWPGLSAYDN  
SPRSPSPSPATPPSQSSCSADMLSTVSHAKNTPSQSHSIQHSRSPSGSVGN  
GSSRYSQNSPTHHIPSRSPSPAKTITPQNAPEERSGRSSFVPEQDQETAKTNGFLK

RTDERSRVLLDRGNIRDKAEKPKGSEKGRADGDWDDQEVLDYFDSKESAKQKPHD  
SEGDDTEEDYQFRKSVLADQKSPATSHRNTSEEGPKYKSVLSKGRSSDGR  
EENYKLKETAYIVERPSTAKDKEEDKSDRIITVKEVQSPQVSEKSELKELFDYS  
PPLHSLDAREKSIIFREESPLRIKMIASDSHRPEVKLMAFVPLDDNSRPAISLTKDRL  
LASTLHVSVKKEQEFPSIPDHILKPOANKSTSESFIOHIVSLVHHVKEQYFKSPAVTL  
NERFTSVOKATEHSTROKSPETHRIDISPSALRKHTRLAGERGPKESIQKGDKKL  
RCSADLRHDIDRRRKERSKERGDSKGSRESSGRKQEKTPKDYKEYTPYKDDSKNVF  
VFYIAYGG"  
polyA\_signal 3145. .3150  
/notes="putative"  
polyA\_site 3166  
/notes="putative"  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.61 Length: 3166  
Score: 144.00 Matches: 137  
Percent Similarity: 32.91% Conservative: 97  
Best Local Similarity: 19.27% Mismatches: 243  
Query Match: 3.46% Indels: 234  
DB: 11 Gaps: 30  
US-09-765-271-56 (1-796) x AK088741 (1-3166)  
QY 121 AlaHisAlaAspAsnValArgThrLysGluLuleAsnArgGlnLysGlnGluHisSer 140  
Db 213 TCACATCTTCAAGATCCAAAGTCTAGATCACACTAGTTCTTAGATCAAGATCAAGATCA 272  
QY 141 GlnHisArgGluGlyThrProAigAsnAspGlyAlaValAlaLeuAlaAgsSerGln 160  
Db 273 CATTCAGAAAGAGAGATACAGGTCTCTCCAGAACATATTCGAGGTCTCTGTAGTAGTA 332  
QY 161 GlyArgTyrThrThrAspAsp  
Db 333 GATCGTATTATTCTAGAGATTATCTCGAGATTACAGGAATAATAGAGAAATGAGACGA 392  
QY 168 -----GlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIle 185  
Db 393 CCTTATGGGTAC-----AGAGGAAGGGGTAGAGGGTATTAT 428  
QY 186 ValProHisGlyAspHisTyrHis-----TyrIleProLysAsnGluLeuSerAla 202  
Db 429 CAAGGAGGAGGAGGAGATACCATCGAGTGGCTATAGACCTCTCTGGAATAGAGGCAC 488  
QY 203 SerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArg 222  
Db 489 TCTAGG-----AGTCTAGACAGGTCTCGGTCCGTCACGGTCCAGG 524  
QY 223 ThrTyrArgArgGlnAsn---SerAspAsnThrSerArgThrAsnTrpValProSerVal 241  
Db 525 AGTCCAAAAGAGAGATCCCGTGTCTTCAAGATCCCGAAGCAGATCTCCCGGTCTATAT 584  
QY 242 SerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAla 261  
Db 585 AGATCTCTAGGTCTCCAGATCAATCATCTCGCTCTTCATCCCATATAGCAATCT 644  
QY 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrIlyLeu----- 277  
Db 645 CCTGTCTCTAAAAGACAGGGTCTCAGGAAAAACAAACCAAAAGCTCAGGGGGGAAACCC 704  
QY 278 -----ProLeuSerGlnArgHisValGlu-----SerAspGlyLeuValPhe 291  
Db 705 CAAGAAGAGAGTCTTTGAAAGCAATCAGAGGAGGAACCAAGAGATATCTTTGAACAT 764  
QY 292 AspProAla-----GlnIleThrSerArgThrAlaArgGlyValAlaValPro 307  
Db 765 GATCCATCTGATCTATTGATGAGTTTAATAATCTGCTACTCTCTGCTGATATTTGGCT 824  
QY 308 HisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeuGluArgIle 327  
Db 825 GGCCTTTCAGCTTATGATAATAGTCCAAGTCA----- 857  
QY 328 AlaArgIleIleProLeuArgTyrArgSerAsnHisTrpValProAspSerArgProGlu 347

```

Db      858  -----|||-----CCTCAT 863
Qy      348  GlnProSerProGlnProThr-----ProGluProSerProGlyProGlnProAlaProAsn 366
Db      864  AGTCCTTACCTATTGCTATACACACCCAGTCAGATTCATCATGCTCAGATCCGCCCATG 923
Qy      367  LeuLysIle-----AspSerAsnSerSerLeuValSerGln 378
Db      924  CTTAGTACAGTGCACTCTGCCAAATAATCCCTTCTCAGCATTTCATCTCAGCATTCAGCAC 983
Qy      379  LeuValAlaGlyLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrVal 398
Db      984  AGTCCTGAGAGGTCTGGATCTGGCTCTGTT-----GGAAATGGTGGTGCATATAGT 1037
Qy      399  PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln 418
Db      1038  CCTTCTCAGATAATAGTCCGATTTCATCATCCATCCATCA-----CGA 1076
Qy      419  GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu 438
Db      1077  AGAGCCCTGCCAGACAAATCACCCACAGAAAT-----GCTCCAGAGAGAGAA--- 1124
Qy      439  PheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLys 458
Db      1125  -----TCTAGG 1130
Qy      459  GlyArgAsnSerAspPheGlnAlaLeuAsp-----LysLeuLeu 471
Db      1131  GGGCGGCTCTCTTTTATCTCTGAAGAGATCAGAAACAGCAAGACAGGAAAGTTTTTA 1190
Qy      472  GluArgLeuAsnAspGluSerThr-----Asn 480
Db      1191  AAAGGTTTCACAGATGAAGAGTCTAGAGTATTCCTGCTGTAGGGTAATATCAGGGAT 1250
Qy      481  LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db      1251  AAAGAG-----GCTCCAGAGAGAAAGGGTCAGAG 1280
Qy      501  LeuGlyLysProAsnSerGln-----IleGluTyrThrGluAspGlu 514
Db      1281  AAAGCAGGGCAGATGGAGATGGGATGACACAGGAAGTTCTAGATTACTTTAGTATAAA 1340
Qy      515  ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu 534
Db      1341  GAGCTCTGCAAAACAA-----AAATCCATGACTCTGAAGGGGATCAGACAGAGGAG 1391
Qy      535  HisAsp-----IleIleSerAspGluGlyAspAlaTyrVal 546
Db      1392  ACAGAGGATTATAGACAGTTTAGGAAATCAGTCTCAGCAGATCAGGGAAGAGCTTTGCT 1451
Qy      547  ThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys 566
Db      1452  ACT-----TTTAGAGAGAGAA 1454
Qy      567  ValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAsp 586
Db      1455  ---TCATCTCACCAGGAACTAGCAGAGGAGGAGGAGCCCAAGTCAAGTCC---AAAGTTTCA 1508
Qy      587  ValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGlu 606
Db      1509  TTTAAAGGCAATAGAGAAAGTATGGA-----TTTAGAGAGAGAA 1547
Qy      607  LysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsn 626
Db      1548  AAAAATTATAACTGAAGAGACTGCATACATAGTTCGAAGCCCTAGCCTCAAGGAC 1607
Qy      627  GlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAsp 646
Db      1608  -----AAGCACAAGGAA 1619
Qy      647  AspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla----- 663

```

```

Db      1620  -----GAAGACAAAGGCTCTGTAGATAATA 1643
Qy      664  ThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTyrPgly 683
Db      1644  ACAGTGAAGAAAGAGTGCAGTCCCGGAGCAG----- 1676
Qy      684  AsnAlaSerGluHisLeuLeuGlyLysAspHisSerGluAspProAsnLysAsnPhe 703
Db      1677  GTAAAGTCTGAAAGCTCAAAGAGCTCTTTTATTACAGTCCCTCTTACACAGAGTCTG 1736
Qy      704  LysAlaAspGluGluProVal-----GluGluThrProAlaGluProGluValProGln 721
Db      1737  GACGCACAGAAAAATCTCATCTTCAGAGGAGAGCCCACTGAGGATCAAAATGATAGCC 1796
Qy      722  ValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysVal 741
Db      1797  AGTGATTTCTACCGTCTCTGAAGTCAATCAAAATGGCACCTGTTCCTCTT----- 1847
Qy      742  ThrAspSerSerLeuLysAlaAsnAlaThrGlu 752
Db      1848  GACGATTCACACAGACCTGCTCTCTGACTAAA 1880

RESULT 6
BC049893      3706 bp      mRNA      linear      HTC 08-APR-2003
LOCUS      Mus musculus, Similar to protein tyrosine phosphatase, non-receptor
DEFINITION      type 3, clone IMAGE:5397303, mRNA.
ACCESSION      BC049893
VERSION      BC049893.1 GI:29612632
KEYWORDS      HTC.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 3706)
AUTHORS      Strausberg,R.
TITLE      Direct Submission
JOURNAL
SOURCE
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Iaric,P., Legaspi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAK Plate: 99 Row: p Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6755239
This clone has the following problem: frame shifted.
Location/Qualifiers
1..3706
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"

FEATURES
source

```



```

/clone="IMAGE:5397303"
/tissue_type="mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"

```

## ORIGIN

## Alignment Scores:

```

Pred. No.:      2.04      Length:      3706
Score:          144.00     Matches:      142
Percent Similarity: 31.96%   Conservative: 75
Best Local Similarity: 20.91% Mismatches: 234
Query Match:      3.46%     Indels:      229
DB:              11        Gaps:       33

```

US-09-765-271-56 (1-796) x BC049893 (1-3706)

```

Qy 148 ProArgAsnAspGlyAlaValAlaLeuAlaAArgSerGlnGlyArgTyrThrThrAspAsp 167
Db 1119 CCTGCAAAAGCGAGCGCGCCGCTGAACCTCAGTCAGGAGGACTC----- 1163
Qy 168 GlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValPro 187
Db 1164 -----ATCCTCTAGGATGTCTCTGACAAACATCTTTGTGCC 1199
Qy 188 HisGlyAspHisTyrHisTyrIleProIlyAsnGluLeuSerAlaSerGluLeuAlaA 207
Db 1200 AACAAAGATGATTCCTACTCTCCAGACAGCCCTGGATAGAACCCAGATTGAC--- 1256
Qy 208 AlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgTyrThrTyr----- 225
Db 1257 -----CTCAGTGTGTCG-----ATCCGCAAGCGCAGCTCTACAGTGCACAC 1298
Qy 226 ArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer-----ValSerAsn 243
Db 1299 AGTACGAATTCCTTAAACACACTTCAGCCCTACCTGCGAGCCCTCCCGCATGCTCCCAAC 1358
Qy 244 ProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263
Db 1359 CCGAGTATCACTGGCAGTGATGTCATGAGACCCGAGTACCTCCGTCCTCCACCGCACAGT 1418
Qy 264 -----SerAsnAspIleAspSerLeuLeuGlyGlnLeu 274
Db 1419 GCGCTCATTCCCGCGTCTTACCGCCCGACCCCTGACTATGAGCCGTGATGAGCAGCTC 1478
Qy 275 TyrIlyS---LeuProLeuSerGlnArgHisValGluSer----- 286
Db 1479 AACAGAGGCATGTTGTCACGACGACCGGACAGCCCTCGCTGAGGAACCTCAACATTGGC 1538
Qy 287 -----AspGlyLeuValPheAspProAlaGlnIleThrSer 298
Db 1539 AGTTCTATCGGTACAGCGCGCGCGCTTGCTCTACAGCCAGCCGCGAGATT----- 1592
Qy 299 ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPhe---IleProTyr 317
Db 1593 CGGAGCACCCTCACCCTCACCTCTCCCGAGTCAGCCCACTACCCCTTTAACCTGAACATC 1652
Qy 318 SerGlnMetSerGlu-LeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSe 337
Db 1653 AGTTTCCACAGCAGTCTC-----CGTATCCCTACCTCCGCGAGAGCGG 1697
Qy 337 rAsnHisTrpValProAspSerArgProGlu-----GlnPr 349
Db 1698 CCAGTGTGGGCGCGTGG-AGTGTGCTGAGCTGACAAATGTGCGAGCTGCAGGCTCAGGA 1756
Qy 349 oSerProGlnProThr-----ProGluProSerProGln 360
Db 1757 CTACCCAGCTCCAAACATTATGAAACCCAGGTGTACCGGCCACCCCGCGCTTACCTTA 1816
Qy 360 yProGlnProAla-----ProAsnLeuLys-----IleAspSerAs 372
Db 1817 CCCAAGACCTGCCAAACAGCACCCCGACACCTGTCCCGCCACCTCTACATCAGCAGCAGCA 1876

```

```

Qy 372 nSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluY 392
Db 1877 TCCGATCTCATCACAGCGCGTCCATCACTCACTGTCAG-----ACCTTCCAGAGGA 1930
Qy 392 sGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGl 412
Db 1931 CAGCTTGGCCCGTGGTCACTCTCTGCGAGGAGGTGAGCGAGCCCTCACAGCAGCGCGCA 1990
Qy 412 uSerLysLeuSerLysGlnGluSerVal----- 421
Db 1991 TGCCACCTACAGAAGAGAGACAGTATTGAATCGCAGGGCTCACACATGATTCGAAGG 2050
Qy 422 -----SerHisThrLeuThrAlaLysLysGluAsnValAlaProArg----- 435
Db 2051 CCTGAGGCTCAAGGAGAGGAGCGTGTCTGAGCCTCAGCAGCAGCGTGGCGCCCGACCTT 2110
Qy 436 -----AspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAl 450
Db 2111 CTCAGCAGGCTCCAGTCCAGCTCCAGCTTCTCTGACAAATGAAGCAGGAGGGGACTGAAGA 2170
Qy 450 aHisLys---AlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLy 469
Db 2171 GCAGGAGGTGGCAGATACAGCCCATAGAAGTCCCTTCTGATGCCACCATGCTGATTCA 2230
Qy 469 sLeuLeuGluArgLeuAsnAsp-----GluSerThrAsnLysGluLysLeuValAspAs 487
Db 2231 TAGCAGTGAGGAGGAGCAGGAGCTTGGAGAGGACAGCAGCAGAGACAGGCGCATCTCAGC 2290
Qy 487 pLeu-----LeuAlaPh 491
Db 2291 TGTGCTGAGCCCGCCTTACGCGCGCTTCTCTCAGGAATAAACTACCCCTGTGCTTC 2350
Qy 491 eLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrTh 511
Db 2351 AGCGACTCCGATCACTCGGCTCTGCTGATATTTTGGAGCCCAAGCCCATGTC-----AC 2404
Qy 511 rGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIl 531
Db 2405 AGAGCTTGAGAAGAGGCAAGGACATCAGCCCTGTCCAC-----CTGGT 2449
Qy 531 ePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGl 551
Db 2450 TGTGGAGACCCAT-----CGGCCCCGAGAGAGATGGAGTCTGCTGACCCCTCCATGTC 2500
Qy 551 yHisSerHisTrpIleGly----- 557
Db 2501 CGAGTCAGACCTCAGCAGCTCAGGAGGTACCGAGCTAGGAGGAGCTCTGTCAAGAAAG 2560
Qy 558 -----LysAspSerLeuSerAspLysGluLysValAla----- 568
Db 2561 GCCAGTGTACAGCTCTCTCTCGGAAAAAAGCGCTGTGGAAGAGCTTCGCGCACTAGG 2620
Qy 569 -----AlaGlnAlaTyrThrLysGluLysGlyIleLeuPro----- 580
Db 2621 GGGGATGAAAAAATCGAGCAGATGCAAAAAAATTTGGTCCCTCAAGCTGGCAGCCCT 2680
Qy 581 -----ProSerProAsp-----AlaAspValLysAlaAs 590
Db 2681 CAATGGGCTCTCCCTGTCCCGGCTGCCCTGCTGATGAGGAAGAAAGAGTGTCCACAG 2740
Qy 590 nProThrGlyAspSerAlaAlaIle----- 599
Db 2741 AGCGACAAATGACAGAGGTGCAAGGTCTCTGAGCAGCGGTAGAGCAAGGATGGTCTT 2800
Qy 600 -----TyrAsnArg-----ValLysGlyGlu-LysArgIlePro- 610
Db 2801 CACCGAATGAGAGGATCCTTAAGAAGCGGTGTGTGATGGCGAGTGTCTGACTGCGCG 2860
Qy 611 -----LeuValArg- 613
Db 2861 GCTCCCTGAAATGCAGAGAGAAATCGATTTCAGATGTGCTCCCTACGATGATGGAG 2920

```

Qy 614 -----LeuProTyMetValGluHisThr----- 621  
 Db 2921 AGTCAGTTGTGTCACACAGGAGAACACACCTGGCTACATCAACGGCTCACACATTAA 2980  
 Qy 622 -----ValGluVallysAsnGlyAsnLeuIleIleProHisLysAspHisTyr 637  
 Db 2981 GGTCTCTGTACGGGAATAGATGGATTATATTGCCACACAGGACCATTAAC 3033

RESULT 7  
 AK003180  
 LOCUS  
 DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1100001J13 product:NULP1 homolog [Mus musculus], full insert sequence.  
 AK003180  
 VERSION AK003180.1 GI:12833685  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 10349636

2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159

3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wakahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861

4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2581)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission

TITLE

## JOURNAL

## COMMENT

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCAACTGAGTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGATCCAGAGCTCAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

## FEATURES

## source

## Location/Qualifiers

1..2581  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:1100001J13"  
 /db\_xref="MG1:1894979"  
 /db\_xref="taxon:10090"  
 /clone="1100001J13"  
 /tissue\_type="whole body"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="18-day embryo"  
 51..1930  
 /note="NULP1 homolog [Mus musculus] (SPTR|Q9ER19, evidence: FASTI, 99%ID, 99.8%length, match=1072) putative"

## misc\_feature

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.72 Length: 2581  
 Score: 142.00 Matches: 175  
 Percent Similarity: 30.92% Conservative: 85  
 Best Local Similarity: 20.81% Mismatches: 272  
 Query Match: 3.41% Indels: 310  
 DB: 11 Gaps: 43

US-09-765-271-56 (1-796) x AK003180 (1-2581)

Qy 12 ValLysGluAsnAsnArgValSerTyrIleAspGlyLysGlnAlaThrGlnLysThrGlu 31  
 Db 216 GTCCGTGTCAACACCGGTTTGAGCTGATAAAC-----ACGGAG 254  
 Qy 32 AsnLeuThrProAspGluValSerLysArgGluGlyIleAsnAlaGluGlnIleValIle 51  
 Db 255 GATCTGTGAGGAGACTTGTG-----GTAAATGGGAGAGACTGACTGT 299  
 Qy 52 LysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyrHisTyrTyrAsnGly 71  
 Db 300 ACCCTGCCGAGTCT-----GTGTCATCAGGGAAC-----AAAGGA 335  
 Qy 72 LysValProTyrAspAlaIleIleSerGluGluLeuLeuMetLysAspProAsnTyrLys 91  
 Db 336 AGGGCT-----AAACATGGAACGACGAG 359  
 Qy 92 LeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysValAspGly 111  
 Db 360 ACCAAGCAAGATGGA-----GGGGCCACCAAGCAGG----- 392  
 Qy 112 LysTyrTyrValTyrIleLysAspAlaIleHisAlaAspAsnValArgThrLysGluGly 131  
 Db 393 -----TCCTCGGAACAGCTTAATGCAAGTGCACAACTCCGA 428



Db 2135 GAA 2137

RESULT 8  
AK085512  
LOCUS  
DEFINITION  
AK085512 2532 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D630036P23 product:TARA homolog [Homo sapiens], full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK085512  
AK085512.1 G1:26102774  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
MEDLINE  
JOURNAL  
PUBMED  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20530913  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20530913  
11076861  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20530913  
11076861  
6  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222).

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.

FEATURES  
source  
1..2532  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM\_DB:D630036P23"  
/db\_xref="MGI:2422620"  
/db\_xref="taxon:10090"  
/clone="D630036P23"  
/tissue\_type="kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
118..1861  
misc\_feature  
/note="TARA homolog [Homo sapiens] (SPT9H2D6, evidence: FASTY, 86.9% ID, 100% length, match=1740)  
putative"  
2511..2516  
polya\_signal  
/note="putative"  
2532  
polya\_site  
/note="putative"

ORIGIN  
Alignment Scores:  
Pred. No.: 2.02 Length: 2532  
Score: 141.00 Matches: 138  
Percent Similarity: 30.38% Conservative: 88  
Best Local Similarity: 18.55% Mismatches: 247  
Query Match: 3.39% Indels: 271  
DB: 11 Gaps: 30  
US-09-765-271-56 (1-796) x AK085512 (1-2532)  
QY 146 GlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThr 165  
Db 64 GGTGGCGCGGAGCGCCCGCGCGGTGTCGTCGAGCGTCCCGGGTCCCGCATGACG 123  
QY 166 AspAspGlyTyrIlePheAsnAlaSer-----AspIleGluAspThrGlyAspAla 183  
Db 124 CCGGATCTGCTCACTTCAAGAAAGGATGGATGTCATCTGGATGAGCTGAGAG--- 180  
QY 184 TyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSer 203  
Db 181 -----TGAAAAAGCATTTGGTTGTGTCGACGGATTCAAGCCTCAATATTATAC 228  
QY 204 GluLeuAlaAlaGluAlaPhe-----LeuSerGlyArgGlyAsnLeuSerAsn--- 220  
Db 229 AGAGACTCCACGGCTGAGGAGGACGATGAGCTGGAGCGGTGAGATCGACCTCGGCTCT 288  
QY 221 -----SerArgThrTyrArgArgGln----- 227  
Db 289 ACTGAGCTCAGAGATATCGGTCGAGCGCACTATGGCTTCAGATCCACACCAAGGAC 348  
QY 228 -----AsnSerAspAsnThrSer-----ArgThrAsnTyrValProSerVal 241  
Db 349 GCGGTTTATACCTGTGGCCCATGACCTCGGCATCCGCGCACTGGAATGAGGCTCTC 408  
QY 242 -----SerAsnProGlyThrThrAsnThrAsnThrSerAsnAsn 254  
Db 409 AGGAAAACCTGTCGCGCAACTTCAGCCCGCAGATGTCCACCAAGCTCTCCGACTGCAATAA 468  
QY 255 SerAsnThr---AsnSerGlnAlaSerGlnSerAsnAspIle----- 267  
Db 469 GAAACACGCTGCATGGCTATGGCACCACCCAGAGAGCTCCCTGAAGATAGGGAGCAGCG 528

QY 268 -----AspSerLeuLeuLysGln 273  
 Db 529 ACAGCTCTGAGGTATTCGGCGGGGTGGTCTCCCAAGGCGGATGGACACCGCGGTCA 588  
 QY 274 LeuTyrLysLeuProLeuSer----- 280  
 Db 589 CTGACTACGTGGAGCTCTCTCCATTAGCGCCAAAGCTCCCCACAGCGCATGGCACTCTC 648  
 QY 281 -----GlnArgHisValGluSerAspGlyLeuValPheAspProLagln 295  
 Db 649 TCCCGCTCCATCCCGAGCGCCCAACCAAGCAGGAGGATCTGGAGCGGGACTTGGCCCAA 708  
 QY 296 IleThrSerArg----- 299  
 Db 709 CGCTCTGAGGAACAGCAAGTGGTTCCAGTCCACGGACGGACGCCACCCAGAGAGCGCC 768  
 QY 300 -----ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIle 315  
 Db 769 AGTGGCGATGGTCTCGAAGGGGCTGGGTGCCCTCCCTGACTGATGACCAAG----- 822  
 QY 316 ProTyrSerGlnMetSer---GluLeuGluGluArgIleAlaArgIleIleProLeuArg 334  
 Db 823 -----ACAGGCTCAGTGAGGAGATTGAGAAGAGTGGCAGGAGTTGAAAAGCTGCC 876  
 QY 335 TyrArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGln----- 352  
 Db 877 CTGGGGAGAACAGCGAGTGCCTACCCGCCCTGCTCAACCAAGCCCAACGATCG 936  
 QY 353 -----ProThr 354  
 Db 937 TCGGGGACCTACAAGTACACCATGAGGCCCTGGAGAAGGAGGTCCAGTCTCTCCGTGC 996  
 QY 355 Pro-----GluPro-SerProGlyProGlnProLapProAsnLeuLysII 369  
 Db 997 CCAGCTGGAGCTTGGGCTCTCCAGGGGAAGCTCTCAGAAATGCACCGACATCCAGGA 1056  
 QY 369 eAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPh 389  
 Db 1057 GGACAGCCAT-----ATCCCCCGGCTACATCTC 1086  
 QY 389 eGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrVal-- 408  
 Db 1087 GCAGGAA---GCCTGTGAGCGCAGCTGGCAGAGATGGAGTCCCTCACACGACGAGTGAT 1143  
 QY 409 -----LysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisTh 424  
 Db 1144 GGACAGCTGCAGCGCGCACCATGACGAGCTGCAGCGGTGCAGCGGCTGCAGGAGAGAGGTG 1203  
 QY 424 rLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTy 444  
 Db 1204 GCTCTGGCGGAGGAGCGGTGCCCGCTCAGCTATTGAACCCATGAAGAGGGGTA 1263  
 QY 444 r-----AsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArgAsnSe 462  
 Db 1264 TCAAGAAGAGCTAAGCCCGGAGCTGACAAACCCCGGAGTCTCCAGCAGCGGCCACGAGAG 1323  
 QY 462 r-----AspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAs 475  
 Db 1324 CCTTCGAAGCAACACCGCTGGAGTGGAGCGCGCTAAAGCGAGGCTGCAGGTGCTATC 1383  
 QY 475 nAspGluSerThrAsnLys----- 481  
 Db 1384 TGAGAGTACTCGCAGAAGTGCCTGGAGATCGGGCCCTGCACAGGCGAGGAGAGCG 1443  
 QY 481 ----- 481  
 Db 1444 TGAGCACACCTCGCGGCTGCCAGCAAGAGGCCAGGAGTGTGCGCCACCAACGAGGA 1503  
 QY 482 -----GluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleTh 496  
 Db 1504 GTTGCACTCCCACTGTCCGAGGAGATCGACCGCTCTCGCAGCTTTCATCGCTCGCAGG 1563  
 QY 496 rHisProGluArgLeuGlyLysProAsn----- 505

Db 1564 CACAGGCACAGCTCGCGGCGCAGCAATGAGAGAGTTCCTGTGTGAGCTGGAGGTCTCTCT 1623  
 QY 506 -----SerGlnIleGluTyrThrGluAspG1 514  
 Db 1624 CGAGTGAAGGAGAATGAACCTCCAGTACCTGAAGAAGGAGGTGCAGTCTCTCCGAGATGA 1683  
 QY 514 uValArgIleAlaGlnLeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspG1 534  
 Db 1684 GCTCAGGTGTATACAGAAGGACAAGCGCTTCACGGGAAGTACCAGGATGTGTCTGTGA 1743  
 QY 534 uHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHi 554  
 Db 1744 GCTGAACACATCAAGACTCGTGGAGCTGAGAGCTGAGCAGCTGAAGAGGACCTCGC 1803  
 QY 554 sTrpIleGlyLysAspSerLeuSerAspLysGluLysVal----- 567  
 Db 1804 C---CTGGCCATGGCTGCCCTGCAGGAGAGGAGGCTGTGAGGAACAGCTTTGGCTGTGA 1860  
 QY 568 -----AlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProPr 581  
 Db 1861 GAGTCCCTCTGGTTCGATCAGCTGAGGAC-----CCTCC 1896  
 QY 581 oSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAs 601  
 Db 1897 AGCCCCA-----GGGACCCAGCCACCTCTCTCTCGG 1929  
 QY 601 nArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisTh 621  
 Db 1930 CAGAAGTCAGAAGCAAGCTTCTCCCC-----ACCCTGGGTCACTGTGCACCTCACAC 1983  
 QY 621 rValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIleLy 641  
 Db 1984 CACACACACC-----CACACACACACG----- 2005  
 QY 641 sPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLe 661  
 Db 2006 -----CACATACACACACACACACACACGATATGCACAC 2046  
 QY 661 uPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspG1 681  
 Db 2047 ACATGCA-----CACACATACACGACACACACATGCACAGACA 2085  
 QY 681 yTrpGlyAsn 684  
 Db 2086 CTGGGGATCT 2095  
 RESULT 9  
 AK029960  
 LOCUS  
 DEFINITION  
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493241D20 product:similar to CTCL TUMOR ANTIGEN SE2-2 (FRAGMENT) (Homo sapiens), full insert sequence.  
 AK029960  
 VERSION  
 AK029960.1 GI:26325891  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 PUBMED  
 10349636  
 REFERENCE  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 MEDLINE

```

11042159
3 Shbata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
  Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
  Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
  Yamamoto, S., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
  Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
  Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
  Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
  RIKEN integrated sequence analysis (RISA) system-384-format
  sequencing pipeline with 384 multicapillary sequencer
  Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
  FANTOM Consortium.
  Functional annotation of a full-length mouse cDNA collection
  Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
  Group Phase I & II Team.
  Analysis of the mouse transcriptome based on functional annotation
  of 60,770 full-length cDNAs
  Nature 420, 563-573 (2002)
6 (bases 1 to 2740)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
  Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
  Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
  Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
  Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
  Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
  Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
  Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
  Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
  Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
  Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
  Muramatsu, M. and Hayashizaki, Y.
  Direct Submission
  Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
  Physical and Chemical Research (RIKEN), Laboratory for Genome
  Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
  RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
  Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
  URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
  Fax: 81-45-503-9216)
  cDNA library was prepared and sequenced in Mouse Genome
  Encyclopedia project of Genome Exploration Research Group in Riken
  Genomic Sciences Center and Genome Science Laboratory in RIKEN.
  Division of Experimental Animal Research in Riken contributed to
  prepare mouse tissues.
  Please visit our web site for further details.
  URL: http://genome.gsc.riken.go.jp/
  URL: http://fantom.gsc.riken.go.jp/
  Location/Qualifiers
    1..2740
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="FANTOM DB:4932411D20"
      /db_xref="MGI:2392537"
      /db_xref="taxon:10090"
      /clone="4932411D20"
      /sex="male"
      /tissue_type="testis"
      /clone_lib="RIKEN full-length enriched mouse cDNA library"
      /dev stage="adult"
      39..2465
        /note="unnamed protein product; putative
        similar to CTCL TUMOR ANTIGEN SE2-2 (FRAGMENT) [Homo
        sapiens] (SPTF|QH266, evidence: PASTY, 83.6%id,
        93.6%length, match=2139)"
        /codon_start=1

```

```

/protein id="BAC26700.1"
/db_xref="GI:26325892"
/translation="MLQTHASEVKKVKAIVEDLRHLAAQAAHQDSQSLKSELAQOKEA
NSRAPTTMRNLVRLKSLQALKKEQKALSRALLESEMTAAAEERIIIAVTSQKEA
NLNVQVVVHRETRKLSQEDENLNLKALKEALTSKNKNSLADDLNELLNKKQKQ
KAYNKLREKDGIDQENDELRRQIKLSSGLQSKTLIDNQSILDELQKVKKLSSQL
ERKVDVDIKPVKEKSEELIRWEGKQWTKVGRNRLKEKEGEAGHJLQNLMLR
KELPAADKELTLQKLUKTTGMTVDQVGLVRALESEKELBELKKKJNDLENDIILM
TQOALPRSVVEDLHLQNKYLQKLUHTLEKLSKESQSLTSEIETSDHCOKEQLQ
KENLKSSENLELFQLEQANLPRLNQVLDKMECFLEKLEKLEKLEKLEKLEKLEK
GRSGKTIPELEKTIIGLMKKLVEKVORENEQLKASGILTSKMATIEENRNKAELE
KLKAHFQSLQSFESKNGKTEKIVAENRERKLEKLEKLEKLEKLEKLEKLEKLEK
MAAQLEETGKQLQFASRAPLEGADSKWSIVSVRYVETKMEKLEKLEKLEKLEK
DLKOLVREATEREQAKKYTDLEQIILANPREGATEQELIRELQLLRLANNQMD
KERALIHQIEINKQDTRADSSIPDSQLEKINDLETKLEKLEKLEKLEKLEKLEK
KELENFDSFPEEIEDLKYNKEEVKKNLLEKLEKLEKLEKLEKLEKLEKLEKLEK
GESPSFFPIY"
ORIGIN
Alignment Scores:
Pred. No.: 2-5 Length: 2740
Score: 140.50 Matches: 118
Percent Similarity: 34.98% Conservative: 97
Best Local Similarity: 19.13% Mismatches: 223
Query Match: 3.37% Indels: 177
DB: 11 Gaps: 24
US-09-765-271-56 (1-796) x AK029960 (1-2740)
QY 249 ThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspLeu 268
Db 12 TCAGAAACCTTCCACACATTTGTGATGATGCTCCAGAAACATCATCCGACGAGTAAAG 71
QY 269 SerLeuLeuLysGlnLeuTyrLysLeu-----ProLeuSerGlnArgHisValGluSer 286
Db 72 AAGTGAAGACGAGGAGTAGAGACTTTAAGCATGCTCTTACGCCACACACACGAGGACTCC 131
QY 287 AspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaVal 306
Db 132 CAGAGTTTAAAGTCTCAACTCCAGGCTCAGAAAGAACCAACTCCAGAGCT-----182
QY 307 ProHisGlyAspHisTyrHisPheLeuProTyrSerGlnMetSerGluLeuGluArg 326
Db 183 -----CCACACCAACCAATGAGGAATCTTGTAGACAGG 215
QY 327 IleAlaArgIleLeuProLeuArgTyrArgSer-----AsnHisTrpValPro 342
Db 216 CTAAGAGCAACTAGCCTTGAAGAGACCAACAAAGCAGCTAGTCAGCCCTGTG 275
QY 343 AspSerArgProGluGlnProSerProGlnProThrProGluProSerProGlyProGln 362
Db 276 GAACCTTCGGTCGGAATGACAGCAGCAGCTGAGGACGTATATCGCTGTAACCTCTCAA 335
QY 363 ProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLys 382
Db 336 AAGAGGCAAAATCTCAAT-----GTTCAACAGTTGTTAGCGC 374
QY 383 ValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAsp 402
Db 375 -----CATACTAGAGAG 386
QY 403 LeuProSerGluThrValLysAsnLeuSerLysLeuSerLysLeuSerGlnGluSerValSer 422
Db 387 CTAAGTCAAA---ATTGAAGATTAAATGAAATCTTTTAAATTTGAAGAGGCTCT 443
QY 423 HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAsp---441
Db 444 AAACAAAGTAGAACCAACAAAGAAATTCAGTACGTGATGATTTAAATGAATTAATAATGAA 503
QY 442 -----LysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***456
Db 504 CTGCAAAAAAGCAAAAGCTTATAAT-----AAATCCTTAGAGAG 545

```

Qy	457	AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAsp	476
Db	546	AAAGATGGAATTGATCAAGAAAAATGATCAACTGACGAAGACAGATATAAAGACTGTCCAGT	605
Qy	477	Glu-----SerThrAsnLysGluLysLeuValAspAspieuLeuAla	490
Db	606	GGACTGCAGACAAACCTTTGATAGATACCAAGCAAAAGTTTAATCGATGAACCT	659
Qy	491	PheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyr	510
Db	660	-----CAAAAGAAAGTTAAATAAACTTGAAGCCAACTGGAAAGA	698
Qy	511	ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr	530
Db	699	AAGGTGGATGACGTAGACATAAAGCCGCTGAGGAAAAAGAGTAGT	743
Qy	531	IlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThr	547
Db	744	-----AAAGAAGAATTAAATTAGTGTGGGAAGAAGGTAAAGATGGCAACCACAAAGTAGAG	799
Qy	548	-----ProHisMetGlyHisSerHisThrIleGlyLysAsp	559
Db	798	GGACTACGAACACAGACTAAAGGAGAGGAAGGCCACGCCCTGGCAACAGCAGCTG	857
Qy	560	-----SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyr	572
Db	858	AATACCTTAAAGGAACCTTTTGGCCAAAGCTGATAAGAGAACTTACTTTCAGAGAAGA	917
Qy	573	ThrLysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThr	592
Db	918	CTGAAACAACAGCAATG-----ACT	938
Qy	593	GlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuVal	612
Db	939	GTTCAGCAGGTTTTAGGAGTG--CGAGCTTTGGNACTCTGAAAAGAGCTTGCAG	989
Qy	613	ArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlePro	632
Db	990	-----GAGCTAAAGAAAAATCTGCAGCTAGAA	1019
Qy	633	HisLysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAla	652
Db	1020	AATGACATATTATACATGAGCACCAGCAGCT-----CTT	1055
Qy	653	ProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGlu	670
Db	1056	CCAGCAGATCTGTGTGGAGACTTACATTTACAAATAAATACCTTCAGAAAAAACTT	1115
Qy	671	HisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeu	690
Db	1116	CATACCTTTAGAAAAAACTTTCAAAGGAGAAATATTCTCAGTCTTTCAGCTTCAGAAATA	1175
Qy	691	GlyLysLysAspHis-----SerGluAspProAsnLysAsnPheLysAlaAsp	706
Db	1176	GAGTCAGATGATCACTCTCAAAAAGAACAGAACTTCAGAGGAAAAATTTGAAGTTGTCA	1235
Qy	707	GluGluProValGlu-----GluThrProAlaGluProGluValProGlnVal	722
Db	1236	TCTGAACAACATCGAGCTGAAATTTCACTTGAACACAGCAATAAAGATTTGCCAAGACTA	1295
Qy	723	GluThrGlu-----LysValGlu	728
Db	1296	AAGAATCAAGTGAAGAAATTTGAAGAAATGTGCAATTTCTTTAAGAAAGGAATACTGAA	1355
Qy	729	AlaGlnLeuLysGluAlaGluVal-----	736
Db	1356	CTTGACGGGAAGCTTGGTTCAGGTCAGAGGGGCTGGTAGAAGTGGGAAGACAAATCCCGAA	1415
Qy	737	-----LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla	750
Db	1416	CTAGAAAAAACCATTTGGGTTAATGAAGAATATTAGTTGAAAGAGTCCAAAGAGAAAT	1472
Qy	751	ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMet	766

Db 1473 ---GAACAATTGAAAAAGGCATCAGGAATACTGACTAGTAGTGAATAATGGCTACTATTGAG 1539  
 Qy 767 ---AspAsnIenSerIleMetAlaGluAlaGluLysIleuLeuAla 780  
 Db 1530 GAAGAAAATGAAAACCTTAAAGGCTGAACTAGAAAAGCTTAAAGCT 1574  
 RESULT 10  
 LOCUS BC014691 3045 bp mRNA linear HTC 19-NOV-2003  
 DEFINITION Mus musculus protein tyrosine phosphatase, non-receptor type 21,  
 mRNA (cdna clone IMAGE:4527498), containing frame-shift errors.  
 ACCESSION BC014691  
 VERSION BC014691.1 GI:15928418  
 KEYWORDS HTC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
 1 (bases 1 to 3045)  
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
 Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
 Scheetz,T.E., Brownstein,M.J., Uesdin,T.B., Toshiyuki,S.,  
 Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J.,  
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
 Vallalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,  
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvachenko,Y.,  
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
 Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,  
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 TITLE Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 REFERENCE 12477932  
 2 (bases 1 to 3045)  
 AUTHORS Strausberg,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 31 Row: 0 Column: 6  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 6755239  
 This clone has the following problem: frame shifted.  
 Location/Qualifiers  
 1  
 10045







```
Db 1352 -----CGAAGAGCCCTGCAAGAACATCACACCACAGAAAT-----GCT 1390
Qy 434 ProArgAspGluGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1391 CCAAGAGAGGAC-----TCTAGGGGGCGTCCCTCT 1420
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473
Db 1421 TTTTATCTCTCGGAGGAGAGATCAGAAACAGCAAGACAGGAAAGTTTGTGAAGG 1480
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuVal-AspAspLeuLeuAlaPheLeuAl 493
Db 1481 TTTACAGATGAA-----CAGTCTAGAGTATTCTCTGTGATAGGGGTAAATACC 1528
Qy 493 aProIleThrHisProGluArgLeuGlyLysProAsn-----SerGlnIleGluTyrTh 511
Db 1529 AGGGATAAGAGCCCAAGAGGAGAAAGGTCCTCGAGAAAGGAGGACGACGAGATGG 1588
Qy 511 rGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrI 531
Db 1589 -GACATCAGAGGTTCTAGATTACTTTAGTATAAGAGTCTGCACAAACAAAGTTC-- 1645
Qy 531 ePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGl 551
Db 1646 -----CATGAT-----TCTCAAGGGGATGAC-----ACAGAGGAGACAGA 1680
Qy 551 yHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys-----ValAl 568
Db 1681 GGATTATAGACATTTAGGAATCTAGTCTCTGCAGATCAGGGAAGAGCTTTGTCTACTTC 1740
Qy 568 aAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAspVally 588
Db 1741 ATCTCACCAGATATCTAGGAGGAGGA-----CCCAAGTACAAGTCCCAAGATTTC 1791
Qy 588 sAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysAr 608
Db 1792 ACTAAAAGGCAGTAGAAGT-----GATGGATTGTAGAGAAAGAAAAA 1836
Qy 608 gileProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyAs 628
Db 1837 TTATAAACTGAAAGACACTGCATATAGTGAAGGCTAGCACTACAAAGAC----- 1891
Qy 628 nLeuIleLeuProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspPhi 648
Db 1892 -----AGCACACAGC----- 1903
Qy 648 sThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla-----ThrI 665
Db 1904 -----GAAGACAAAGGCTCTGACAGAAATACAGT 1932
Qy 665 eLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTyrGlyAsnAl 685
Db 1933 GAAGAAAGAGTGCGACTCCCGAGCAG-----GTAAA 1965
Qy 685 aSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPheLysAl 705
Db 1966 GTCTGAAAGACTCAAGAGCTCTTTGATTACAGTCCCTCTACACAGAGTCTGACGCG 2025
Qy 705 aAspGluGluProVal-----GluGluThrProAlaGluProGluValProGlnValGl 723
Db 2026 ACCAGAAAATGCCATCTTCAGAGAGGAGAGCCCGCTGAGATCAAAATGATACCCAGTGA 2085
Qy 723 uThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAs 743
Db 2086 TTCTCAGCGTCCTGAAGTCAAACTCAAAATGGCCCTGTCCTCTG-----GATGA 2136
Qy 743 pSerSerLeuLysAlaAsnAlaThrGlu 752
Db 2137 CTCACACAGACCTGCCTCCCTTGACTAAA 2164
RESULT 12
AY404467 3573 bp DNA linear GSS 12-DEC-2003
LOCUS
```

```
DEFINITION Pan troglodytes HCM1900 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY404467
VERSION AY404467.1 GI:39760444
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 3573)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 3573)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarimal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
location/Qualifiers
1..3573
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>3573
/locus_tag="HCM1900"
gene
ORIGIN
Alignment Scores:
Pred. No.: 4.95 Length: 3573
Score: 139.00 Matches: 182
Percent Similarity: 34.57% Conservative: 136
Best Local Similarity: 19.78% Mismatches: 350
Query Match: 3.34% Indels: 254
DB: 29 Gaps: 40
US-09-765-271-56 (1-796) x AY404467 (1-3573)
Qy 31 GluAsnLeuThrProAspGluValSerLysArgGluGlyLeuAsnAlaGluGlnIleVal 50
Db 7 GAGGAGGATACCAAGAGAGAGTCCCAAGAGGATGGAGTT----- 48
Qy 51 IleLysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyrHisTyrTyrAsn 70
Db 49 -----GGTGATGCAACATTTCGNNNC 72
Qy 71 GlyLysValProTyrAsp---AlaIleIleSerGluGluLeuLeuMetLysAspProAsn 89
Db 73 AGCAGATATTCGAGAGCATAGCTACATACAAAGAGGAGTATTCACAA-----AAT 126
Qy 90 TyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysVal 109
Db 127 GTAGCTTAGAGAGAGACTCTGTGGTGCAATCTTCAGTCTCAGATTTAAAGCAATGAGACTAT 186
Qy 110 ASPGly-----LysTyrTyrValTyrLeuLysAspAlaAlaHisLysAspAsn 125
Db 187 GAGAGCCTCAGCCCAAGAAATACTCTTCAGTCTCAGATTTAAAGCAATGAGACTAT 246
Qy 126 ValArgThrLysGluGluIleAsnArgGln-----LysGlnGluHisSerGlnHisArgGlu 144
Db 247 TCTGGCTCTCTATTATTAACCTCTCACACCTGTGCCTGCAAGAGAGAGCCCTTCTCCTTTC 306
Qy 145 GlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGly---ArgTyr 163
```

Db 307 GGAACCAAGTGTGAGAAGC-----AGCCCTCTCTTAGCCAAAATCCAGGAGCAAGCAG 360  
 QY 164 ThrThrAspAspGlyTyrIlePheAsn-----AlaSerAspIleIleGlu 178  
 Db 361 TTCTCTTATGATGGAACATGACAAAATATTCCTATGAAACCTTCAAACTCTTACTCAGC 420  
 QY 179 AspThrGly-----AspAlaTyrIleValProHisGlyAspHisTyrHisTyrIlePro 196  
 Db 421 ACTACATCTCTCCCTCAGTGGATATCCACTTGAAGAGCAGACTTTGATCATTTACTGGC 480  
 QY 197 LysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGly 216  
 Db 481 CGGGAC-----AGTCAAGGGCCCTG----- 501  
 QY 217 AsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsn 236  
 Db 502 AGGCTCTCAGAGAAGCTCCCTATTCCAAATATAGTCAAGGAATAAATCGACGACAAAT 561  
 QY 237 ----- 561  
 Db 562 GTCTACTTTCTTGGAGGCTGGAAGTCCGAAGGCATCTGCTCGCTCGCTCGCCATGTGG 621  
 QY 238 -----ValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253  
 Db 622 AATGGAAGTTCCTCGAGTGTGCTGGCCGCTCTCTATCAGCAGATCGGAGCGCAAGC 681  
 QY 254 AsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln 273  
 Db 682 ATGCTTTCAAGCCCAAGCAGCAGGAAATGAGCATTCAGGACAGCTCAGCGCTTCAA 741  
 QY 274 LeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspPro 293  
 Db 742 -----CCCAAGTTAACTAGACACAGGAG-----CTTGCATCTGAAAC 780  
 QY 294 AlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHis 313  
 Db 781 ATCAATTTGAGAAGTAGAAGTACTCCAGCAGCAGCCTGAGTCACATGGGAGCCTAC--- 837  
 QY 314 PheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleIleProLeu 333  
 Db 838 -----AGCCGATCACTTCCCGAGG 855  
 QY 334 ArgTyrArgSer-----AsnHisTyrValProAspSerArgProGluGlnProSerPro 351  
 Db 856 TTGTACAGAGCCACAGAGAACCCAGCTCTCAGCTTGTCTTCAAGAAATCTCTCTG 915  
 QY 352 GlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAsp--- 370  
 Db 916 GGCAATTTCCAAACGAACAACTTTGGGAAAGGATCTACCTCATAGCGTAATAGACAAT 975  
 QY 371 -----SerAsnSerSerLeuValSerGlnLeuValArg 381  
 Db 976 GACAAATTACCTTAATTTTCTTCTTGTGAGCTCAGGGCTTTTACCCTATAAAACCTCTGCT 1035  
 QY 382 LysValGluGluGlyTyrValPheGluGluLysGlyIle-----SerArgTyrValPhe 399  
 Db 1036 TCTGAAGGCAATCTCTTATGATGTTCTACCTCAGTCTCAGTCTCCAGCTCCAGAGTGGCT 1095  
 QY 400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
 Db 1096 CGGAAGATGCTTCTGGGCTCCACCTCTCTGCTGCTCTGATGACTTTGTATGGGCTTCA 1155  
 QY 420 SerVal-----SerHisThrLeuThr-----Ala 427  
 Db 1156 TATGTGGGGCAAAACCCGAGTCAATCAATTTCTGCTGGGGAGTCAGACAGATTTTGTGG 1215  
 QY 428 LysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeu 447  
 Db 1216 ACCAGGAGGAATCTCTCTTGTGGATCTGTGGAATTTGATGAGGCTGATTTGGAAAGCCTC 1275  
 QY 448 ThrGluAla-----HisLysAlaLeuPhe\*\*\*AsnLysGlyLysArgAsnSerAsp 463  
 Db 1276 AGCAGGCTCAGGAACCCCGCCCTGCTCGGGAACGGAAAGCAGTATTATGCTCC 1335

QY 464 PheGlnAlaLeuAspLysLeuLeu-----GluArgLeuAsn 475  
 Db 1336 ATTTCCAGCAGCTGTGACCTGATGATTATCACCAGCGGAGAGGAGGAGAAAGACTCAGG 1395  
 QY 476 AspGluSerThrAsnLys-----GluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
 Db 1396 GAGCAGGAATGAGCGGATTTGGAGACAGCGCTTGGAGACCATCTCAGTCTCTGTCTCT 1455  
 QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGlu-----Tyr 510  
 Db 1456 GAATACACA-----AAGCTGACCTCGCTTATCTACTGGGACCAACC 1497  
 QY 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530  
 Db 1498 GTGGAAGATGTGCAAAAATCAACAAGAGGCTTGAAGAGCTGCAGCTCTCTGTAGGAG 1557  
 QY 531 IlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMet 550  
 Db 1558 TCTGTGTTTGAAGAGCCCTCATGAGCCCTGAC-----ACAAGATACAGG 1602  
 QY 551 GlyHisSerHisTyrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAla--G 570  
 Db 1603 TGCCACCGG-----AAAGACTCCCTCCCTGATGCAGACTTGCAGAGCTGTGGG 1650  
 QY 570 InAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAsp-----ValL 588  
 Db 1651 AGTCTCAGTTCAGACAGTGCAGCTCTTTTACCCCGCAGGAGCAGCAGGAATGATGAAC 1710  
 QY 588 ysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyGluLysA 608  
 Db 1711 CTGAGCCACTCACCGGACTCTCCACCACTCTCCACTTCCGAAAGCTTCCAGC 1770  
 QY 608 rgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGlyA 628  
 Db 1771 GAGT-CCTTATCTAAGTATCTTACCAAGACCCAGAGGTTATAGTGAAGAACAG 1829  
 QY 628 snLeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAsp 648  
 Db 1830 ATCTC-----AGGAGTTGGCTGCAA--- 1849  
 QY 648 isThrTyrLysAlaProAsnGlyTyr-----ThrLeuGluAspLeuPheAlaThrI 665  
 Db 1850 -----TGAAGAAACCCGGATA-GTCAATCTGAACAACCTCGAGGAACCTTAAGCAAAA 1903  
 QY 665 leLysTyrTyrValGluHisProAspGluArg----- 675  
 Db 1904 TCAAGACATAAATGATCAGATGATGAGTCTTTCAGAGAGTTGGATATGGAATGTGCTC 1963  
 QY 676 ----- 676  
 Db 1964 TTTTGGATGAGAACAGAAAATCTGAAACAACCTTATCAAGGAGAGAGATTTTGG 2023  
 QY 676 roHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHis 696  
 Db 2024 ATCATCTAAACCCGAAATAGCTGAACCTGGAAGAACATTTGTTGGTGAAGAACCAAG 2083  
 QY 696 erGluAsp-----ProAsnLysAsnPheLysAlaAspGluGluProValGluG 712  
 Db 2084 ATGCTGACCTGTTGGATGTTGAAAGCAACACATTTGAA----- 2121  
 QY 712 luThrProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeu 732  
 Db 2122 -----GACCTGGAGTTCCAGCAGCTTCAACATGAGAGCCGCTTAGATGAAGAAA 2170  
 QY 732 ysGlu-----AlaGluValLeuLeuAlaLysValThrAsp----- 743  
 Db 2171 AGGAGAACTTGACTCAACAGCTCTCGTGAAGTGTCTGAATATCAACGGAAACATCGTTT 2230  
 QY 744 -----SerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeu- 757  
 Db 2231 CTAGAAAGGAAAAAATTTCTGCAATGGAAGCAAGCAACCAATCATATTGTTTCAGCAGCTC 2290

```

QY 758 -----ArgAsnAsnLeuThrLeuGlnIleMetAspA 768
DB 2291 AGAGAGAACAGATCATTTTGTGAAGAAAGAAATTAATTAATGATGTTGCAAGAG 2350
QY 768 snAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer----- 785
DB 2351 AAAGAGAGATCTTTGTAATTTGGAAAGAAATACTCCAGCCTCTCTGGGGGAAAGGT 2410
QY 786 -----AsnProSerSer-----ValSerLysGlnLysIleAsn 796
DB 2411 TTCCCGTTAAACCCCAATACTTTTAAAGAGGGGTATATCATGTTAAATGAGATTAAT 2466

RESULT 13
CG756440/c
LOCUS
DEFINITION
P051-3-G10.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION
CG756440
VERSION
CG756440.1 GI:37984016
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus
ORGANISM
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
1 (bases 1 to 1831)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
JOURNAL
MEDLINE
22835951
PUBMED
12884007
COMMENT
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES
Location/Qualifiers
1..1831
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Alignment Scores:
Pred. No.: 1.98 Length: 1831
Score: 138.50 Matches: 135
Percent Similarity: 31.09% Conservative: 87
Best local Similarity: 18.91% Mismatches: 239
Query Match: 3.33% Indels: 257
DB: 29 Gaps: 28

US-09-765-271-56 (1-796) x CG756440 (1-1831)
QY 133 AsnArgGlnLysGlnGlnHisSerGlnHisArgGluGlyThrPro----- 148
DB 1765 AACGCACACACCGCAGACACGCCCTCCCGACCGCAGCGCGCGCCGCGCACACCG 1706
QY 149 -----ArgAsnAspGlyAlaValAlaLeuAla 157
DB 1705 CCGCGCGCGCGCCGCCACACCCACCGTCACCAACACACACACCAACCACTAGGTT 1646
QY 158 ArgSerGlnGlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIle 177
DB 1645 CGCGCGCAACCGCAGCAGCAGCAGCAGCAGCAACAAACNATACCAAGCTACTACGCTCAT 1586

```

```

QY 178 GluAsp-----ThrGlyAspAlaTyr-----IleValPro-----HisGly 189
DB 1585 TRATCAATAGCAACAGGACAGCAGCAGCAACCGATACATACCGCCGCGAGACGACGCT 1526
QY 190 AspHis-----TyrHisTyrIlePProLysAsnGlnLeuSerAla 202
DB 1525 AAACACGCGCACCACAAAAAAGCGCGCACACACAGAACACATAACAA-ATAGAGACC 1467
QY 203 SerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArg 222
DB 1466 GAAAAACCTTAGGACAGCAGC-----CGACCACTCGCAACAAACACCA 1422
QY 223 ThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSer 242
DB 1421 AACACGCGCAGCAGCAACAGCGAGACTAGCAGAAC-ACAAATNCCCACGCGCGGTACGA 1363
QY 243 AsnProGlyThrThrAsnThrAsnThrSerAsn-AsnSerAsnThrAsnSerGlnAlaSe 262
DB 1362 ACAAGGCCAACGCCCAACACGAAACACAAACCGCAGCGCACACGCAACAAAGTAGA 1303
QY 262 rGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnAr 282
DB 1302 AAAGCACAAAGAG-----AGAAAGAGCAGACACAAAAGACAGAAAAAACACCA 1255
QY 282 GHIS-----ValGluSerAspGlyLe 289
DB 1254 GCACAAACACACGACACACATCAACCATAGAAACACACCGCAGCAGCAACAAAC 1195
QY 289 uValPheAspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGl 309
DB 1194 AGCGCCACCCACACACAGCAGCAGTGCGCC----- 1161
QY 309 yAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaAr 329
DB 1160 -----CCAACGCACACACCGCGCAACCGCAACCAACCAACCAACCAAG 1120
QY 329 gIleIleProLeuArgTyrArgSerAsnHisTrpValProAspSerArgProGluGlnPr 349
DB 1119 AGAGCACCACCCACACACAGACAG-----CACACACGAACAAAAACGACGACACCC 1066
QY 349 oSerProGlnProThrProGluProSerPro-----GlyProGlnProAlaProAs 366
DB 1065 ACACCAACGACAAAAAACACACACACCCCGCCCATACAAACCGGACCGCCCAATGCCACA 1008
QY 366 nLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGl 386
DB 1007 -----GCATATCCACACAGCAGTGGCACCACAC----- 978
QY 386 yTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGl 406
DB 977 -----ACACAAGAGACACACGACCC 958
QY 406 uThrValLysAsnLeuGluSerLys-LeuSerLysGlnGluSerVal----- 421
DB 957 CACCAACAAAGAACACGAAACACAGCGCGCAGCAGCGGACCGGACCGGACCGCACAC 898
QY 422 -----SerHisThrLeuThrAlaL 428
DB 897 GACCCAAACACCCACCGCCCAAAACACTCNCCCACACACACACACACACACCCACAC 838
QY 428 ysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuT 448
DB 837 GCACACCCCAATCAGCGCGCGCCACACCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGC 778
QY 448 hrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuA 468
DB 777 CGCACGCGCACACCGGCA----- 761
QY 468 spLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspL 488
DB 760 -----CAACACACACACAGAGGAAAAACA----- 737

```

QY 488 euLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnI 508  
 Db 736 -----  
 QY 508 leGluTyThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyThrThrSera 528  
 Db 720 GCGAAGAGCCCAACCAAGATACGA-----  
 QY 528 spGlyTyIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyValThrP 548  
 Db 694 -----  
 QY 548 roHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysVala 568  
 Db 665 -----  
 QY 568 laAlaGlnAlaTyThrLysGluLysGlyIleLeuProProSerProAsp-----AlaA 586  
 Db 664 -----  
 QY 586 sp-----ValLysAlaAsnProThrGlyAspSerAlaAlaA 598  
 Db 619 ATACACCAATCGCGTCCACCTCCATCAGCCACCCGTCGCGGAGACACACCAAG 560  
 QY 598 laIleTyAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyMetV 618  
 Db 559 AACAAAAAAGCAGACAAAGTAAAAAAGAGATTCCGAGCAGCGCG-----GTAA 509  
 QY 618 alGluHisThrValGluValLysGlnGlyAsnLeuIleIleProHisLysAspHisTyH 638  
 Db 508 CCACACCAAGATAGACAAAGACCCACCAAAAT-----CAACAGAACCCAGCACTAAC 458  
 QY 638 isAsnIleLysPheAlaTrpPheAspAspHisThrTyThrLysAlaProAsnGlyTyTyr--T 657  
 Db 457 ACACCCAGCACTAGGACACACACACACACACACACACACCCACCCACCCACACACAGAA 398  
 QY 657 hrLeuGluAspLeuPheAlaThrIleTyTyThrValGluHisProAspGluArgProH 677  
 Db 397 CCCAACACCCCTAACCAACCCCAACCAACCCGCGCGGACACCCC-----CACCCAC 344  
 QY 677 isSerAsnAsp-----  
 Db 343 ACAGCTCAGACGCCCAACCGCCCAACCGCAATTAACACACGCCCCCGGAGGCGCCA 284  
 QY 681 LyTpGlyAsnAlaSerGluHisValLysGlyLysLysAspHisSerGluAspProAsnL 701  
 Db 283 GACAAACAAAGACCAAGACCCCTAATTAATAGAGCAACACACACGACACGACGACGAC 225  
 QY 701 ysAsnPhelys-----AlaAspGluGluProValGluGluThrProAlaGluP 717  
 Db 224 AGAAACACACGACCCCTACACACGACGACCAACACACCCCGGAGACAA---GACGCTGACA 168  
 QY 717 roGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValL 737  
 Db 167 AAGAAATTTTCAAGGGAAGAAACACACACACCAACCAACCCCACTACTACAA----- 118  
 QY 737 euLeuAlaLysValThrAspSerSerLeuLysAlaAsn 749  
 Db 117 -----CAAACTTAAGGGGAAC 100

RESULT 14  
 AK029864  
 LOCUS  
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4931417E02 product:retinitis pigmentosa GTPase regulator, full insert sequence.  
 ACCSSION AK029864.1 GI:26325789  
 VERSION AK029864  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## J

```

/db_xref="MGI:2392424"
/db_xref="taxon:10090"
/clone="4931417E02"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
229..3147

```

CDS

```

/note="unnamed protein product; putative
retininitis pigmentosa GTPase regulator (MGI:1344037,
GB|NM_011285, evidence: BLASTN, 99%, match=1957)"
/codon_start=1
/protein_id="BAC26649.1"
/db_xref="GI:26325790"
/translation="MPGRSRWSQGVGHLRLNRVAPALFPKQAOIPFAGQMAESES
LVPDGVAFVFTKTPFAENIPKFWKNDIPICLSCGDEHTAIVTGNKLMYFGNNW
GQGLGSAKAAIKPTCIKALKEPKVLAACGRNHTLVSTDTGVYAAGNNEGOLGLG
TDDRDTHQIVFPTPADTIKQLSAGANTSAALTDGKLFMMGDSGQIGLEKSNV
CIPHEVTVQVCGGHTVTLTKVVAVFGLGQFGLGTFLFETSEPKIIRIKDQK
LGIIPERVIOVAGGHTVTLTKVVAVFGLGQFGLGTFLFETSEPKIIRIKDQK
ICHISGHNHTALMTELGLLTYFGDGRHGKGLGCMENFTNOFPPTLCSNLFRAFVOLI
ACGGCHMLVATPRTIDEPKFDVIEPIYITSGFSINDLSPRSLSRSLRARRR
ERERPPCSASMTPLPLEGTSASTSAVYPPSPFPHLSVNNYKPSSEMEPLDSD
YFEDKXNDKTETENSADVSENGFTNDILNMTTSSNEKLLDFGPIQKQOQDVT
PEKVMSTPCTENEDSYEYEMSKIKEVTYKQYLAAGIYMRPAEILAEFDSRQVN
GLDQVEERVFDTGKLGQSKQVKSDEIYSEKKEVMEVADVKKIRESESRKSDS
LFDLPDKTMSSESDNKDIABERSSEQNMTFDETELVEPDSYMECERHSEQDSA
LELEQPKLVYESSEKDEKDDDEVENTENLWDRNCTEQETENVFAIRFPFRLD
KHDLSGIPPEEGDEGSENVVQVQAQENLEFEGDRKEAKAPSDVITKEBA
POLSETVPEKEGEMBEISILNVEDTVEERKEGEKEIVEGSIPETEGSETIDITDE
KLDELVBKEDSASLLQRLAREYNENPKMHYDRVKSSSEILGNDPTSKDIKKAKKS

```

polyA\_signal

```

3291..3296
/note="putative"

```

polyA\_site

```

3313
/note="putative"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 5.33 Length: 3313
Score: 138.00 Matches: 131
Percent Similarity: 33.29% Conservative: 107
Best Local Similarity: 18.32% Mismatches: 248
Query Match: 3.31% Indels: 229
DB: 11 Gaps: 28

```

US-09-765-271-56 (1-796) x AK029864 (1-3313)

QY	108	LysValAspGlyLysTyrValTyrLeuLysAspAlaAlaHisAlaAspAsn---Val	126
DB	1468	MAATTGAACACGTATATGACCTTATATAGTACAGGTCTTTTCCATCAATGACCTC	1527
QY	127	ArgThrGluGluLeuLeuAsnArgGlnLysGlnGluHisArgGluGlyGly	146
DB	1528	TCCCCAAGAAGTTCACCTGAATAGATCTTATACGACAGTCTGCGGCGAAGAGCGGAG	1587
QY	147	ThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrAsp	166
DB	1588	AGACCCCCCATGCTCAGCTTCAATGGTG-----GGAACACTGCCTCCATTA	1632
QY	167	AspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVal	186
DB	1633	GAGGGGACTTCTGCCCTCCACTTCA-----GCTATTTTTC	1668
QY	187	ProHisGlyAspHisTyrHis-----TyrIleProLysAsnGluLeuSer	201
DB	1669	CCCAAGTTCACCCCTTCCATTTGCTGTGAATAACTACCCAGAGAAAGCCCTCTGAA	1728
QY	202	AlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSer	221
DB	1729	TCAATGAGCCACTGGACTCAGATATTTTGAAGATAAATGAACAAGACACAGACACA	1788
QY	222	ArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerVal	241

DB	1789	GAAATTCTTCAGCAGTGGATTACAAAACTTTTGGTGAACCTTAATATCTTAATATG	1848
QY	242	SerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAla	261
DB	1849	ACACATATGATGACTACGAGTCCCAATGAGAGTTATTAGATTTTCCCAATTCACAAA	1908
QY	262	SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGln	281
DB	1909	CAACAGATCAA-----GACACATTTGAGAAAGTGAGAAAGTACACCGTGCACTGAA	1962
QY	282	ArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAla	301
DB	1963	AATGAG-----	1968
QY	302	ArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSer	321
DB	1969	-----GATAGTTAAGAA-----TATGAAAGAAATGTCA	1995
QY	322	GluLeuGluGlu-----ArgIleAlaArgIleIleProLeuArgTyr	335
DB	1996	AAAAATAAAGAGTGCACAGTGACAAACAATATTTAGCAAAAGGATCTACATG-----	2049
QY	336	ArgSerAsnHisTrpValProAspSerArgProGluGlnProSerProGlnProThrPro	355
DB	2049	-----	2049
QY	356	GluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeu	375
DB	2050	-----ATACGACACCCGAGATCTGGAAGCATTTTCAGATGAGNA-----	2091
QY	376	ValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSer	395
DB	2092	GTGGGTAATGGCTAGACCAGTGGAGGCGCGGTCTTCTCCTGATGGA-----	2142
QY	396	ArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeu	415
DB	2143	-----AAGGGCTTACAAGCAAGCA-----	2163
QY	416	SerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArg	435
DB	2164	GTAGGAAGAAAGTATGATGAAGAAATAGTCTCTGAGAAAGAACTCAGGTG-----	2214
QY	436	AspGlnGluPheTyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe	455
DB	2214	-----	2214
QY	456	***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsn	475
DB	2215	-----ATGGAAGTGGCAGATGTGAAAGAGATTAGAGAAAGT	2250
QY	476	AspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIle	495
DB	2251	GAGAGAACTCAAAATCAGATTCACCTTTTGTGACTTA-----	2289
QY	496	ThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluVal	515
DB	2290	-----CCAGATAAA-----ACCATCAATTTCTGAG-----AGTGAAGCAATATA	2328
QY	516	ArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis	535
DB	2329	GATATTGCTGAG-----GAAAGGAGAGCAGTGAGCAGATATGACCTTTGACGTGAA	2382
QY	536	AspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrp	555
DB	2383	ACAGAAATGGTAGAAGAACACAGACAGTACATGGAATGTGAGAGG---CACACTGAG---	2436
QY	556	IleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGlu	575
DB	2437	-----CAAGACAGTCTGAGGAGTGGAGCACCCCAAGCTAGTAGAATATAGTAGGAA	2490
QY	576	LysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAspSer	595





Db 2003 -----GAAGCCATCTTCAGGAGGAGGAGGCTGGCCATGTCATGAGGTAGATCTA 2050

Similarity: 32.15%

Similarity: 32.15%



```

QY      747 LysAlaAsnAlaThrGlu 752
Db      1684 CTGCTTCTTGACTAA 1701

RESULT 17
LOCUS   BC048778
DEFINITION Mus musculus, similar to paternally expressed 3, clone
IMAGE:6308481, mRNA.
ACCESSION BC048778
VERSION   BC048778.1 GI:28981350
KEYWORDS Mus musculus (house mouse)
SOURCE   Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 5304)
Direct Submission
Submitted (14-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: Reggen, Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@hri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoev,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 100 Row: B Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6679274
This clone has the following problem: frame shifted.

FEATURES
source
1..5304
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6308481"
/tissue_type="Olfactory epithelium, neonatal mouse,
C57Bl/6J"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 13.1 Length: 5304
Score: 137.00 Matches: 165
Percent Similarity: 36.33% Conservative: 122
Best Local Similarity: 20.89% Mismatches: 273
Query Match: 3.29% Indels: 233
DB: 11 Gaps: 42

US-09-765-271-56 (1-796) x BC048778 (1-5304)

QY      25 GlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGluGlyIle 44

```

```

Db      1118 GAGGCAACTCAGATCAGAGATAATGAGC-----GTG 1150
QY      45 AsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHisGlyAsp 64
Db      1151 AACGGAAACGTGATCGTC-----TACGTG----- 1174
QY      65 HisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGluLeu--- 83
Db      1175 CACGTGCAGAGCAGCGTGCAGCGACGTCGACGGAGCGTGAGCGT-CAGCTTGGG 1233
QY      84 -----LeuMetLysAspProAsnTyr-----LysLeuLysGluAspIle 97
Db      1234 GAACCTTTCTGACCTGTCCTCAAACTTCAATGAGTTTGGGAAGATGTACAGGAACACAAA 1293
QY      98 ValAsnGluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeu 117
Db      1294 ATC-----TATGAGTCAAAAGTGTGGGGAGAGCTTTCTCATCTC 1335
QY      118 LysAspAlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGln 137
Db      1336 TCATCCCTGAGGAGCATCAGAAATCCATACCTAGAGGA-----AACCAATTGAAAT 1389
QY      138 GluHisSerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAla 157
Db      1390 AAGAGCAGGATGTGCGAGGACACTTTGTCCCT-----AGTCAGTCTCTCCGA 1437
QY      158 ArgSerGlnGlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIle 177
Db      1438 CGGCCCAAGAAACCTTACAGAGAGAGCTGTCGACTTTAAACATGCCAGGATGCACCT 1497
QY      178 GluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLys 197
Db      1498 ATGGCAACTCAGACTCC-----AGCGACAT-----CAGAAA 1530
QY      198 AsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGly---- 216
Db      1531 AACCGT-----TCCCGAAGGAACCTTCTTCTGAGGCGCAGAGATTT 1569
QY      217 -----AsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSer 229
Db      1570 GAGAAACCCCTCGTTGAATCTCAGAGAGTCTATCTATTAACAGACCACCTGAAACAAA 1629
QY      230 AspAsn-ThrSerArgThrAsnTyrValProSerValSerAsnProGlyThrThrAsnTh 249
Db      1630 GAGCATGACAAGCCGTTCA-----CCATCAGTGTCA-----ACCCT 1665
QY      249 rAsnThrSerAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSe 269
Db      1666 AATGACAGCTGAAATTCCTCCATCATGGAATAATGGCTCCCGAGGCAAAATCCTATGAGAG 1725
QY      269 rLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLe 289
Db      1726 TCTGTTATT-----CATAGCTTGGGCTCCGAGGAAGCTCAGAGAGTCATGTTGAGCT 1778
QY      289 uValPheAspPro-----AlaGlnIleThrSerArgThrAlaArgGlyValAl 305
Db      1779 GGGGTTTCAGTAAACCAAGACCAAGTGGCAGAGCTAGCAGCCAGAGCTCAAGCAGCATTTA 1838
QY      305 aValProHis-----GlyAspHisTyrHisPheIleProTyrSerGlnMetSe 321
Db      1839 CTACCCAGAGCAGCAGCTCTGGAGGCAACACCTAT-----GAAGAAA 1880
QY      321 rGluLeuGluGluArgIleAlaArgIleIleProLeu-----ArgTyrAr 336
Db      1881 AGATACAGGAGCTCTATCATCCATAGCTTGCAGCTCTCGACCTCTGAAACGTCTAG 1940
QY      336 gSerAsnHis-----TTPValProAsp--- 343
Db      1941 AGCAATGACCATATTCAATGTGATGAGGGGAGAAATCTCCATTATATCCAGATAT 2000
QY      344 -----SerArgProGluGlnProSerProGlnProThrProGluProSerProGlyPr 361

```

Qy	704	ysalaAaspGluGluProValGluGluThrProAlaGluProGluValProGlnValGluT	724
Db	2849	CCATTGATCAGGAGATCGCAGTGAAGAGCCCAAGAGTCTCATGGTCAATGAGC	2908
Qy	724	hrGluTysValGluAAlaGlnLeuLysGluAlaGluValLeuAlaLysValThrAspS	744
Db	2909	CCCATGTTGAAGAGTCCCATGGCAGAG-----AAAGTTGAAGATG	2950
Qy	744	erSerLeuLysAlaAsnAlaThrGlu	752
Db	2951	CTACCATTCAGGCTCATGTTTCTGAA	2976

  

RESULT 18	CG756344/c
LOCUS	CG756344
DEFINITION	P051-3-E10-ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION	CG756344
VERSION	CG756344.1
KEYWORDS	GI:37983823
SOURCE	GSS.
ORGANISM	Pristionchus pacificus
REFERENCE	Pristionchus pacificus
AUTHORS	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
TITLE	1 (bases 1 to 2136)
JOURNAL	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
MEDLINE	An integrated physical and genetic map of the nematode Pristionchus pacificus
PUBMED	Mol. Genet. Genomics 269 (5), 715-722 (2003)
COMMENT	22835951 12884007 Contact: Sommer RJ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel: 00497071601371 Fax: 00497071601498 Email: ralf.sommer@tuebingen.mpg.de Class: BAC ends.

  

FEATURES	Location/Qualifiers
source	1..2136
	/organism="Pristionchus pacificus"
	/mol_type="genomic DNA"
	/strain="California"
	/db_xref="taxon:54126"
	/clone_lib="Ppa EcoRI BAC Library"
	/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

  

ORIGIN			
Alignment Scores:			
Pred. No.:	4.82	Length:	2136
Score:	135.00	Matches:	108
Percent Similarity:	27.13%	Conservative:	67
Best Local Similarity:	16.74%	Mismatches:	212
Query Match:	3.24%	Indels:	258
DB:	29	Gaps:	23

  

US-09-765-271-56 (1-796) x CG756344 (1-2136)			
Qy	235	ThrAsnTrpValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn	254
Db	2088	ACGAAC---GTTCAATCATCAAGTAAACCAACACACACACAAATCGTAACACACACATCAC	2032
Qy	255	SerAsn-----ThrAsnSerGlnAlaSerGlnSerAsnAspIle	267
Db	2031	CAGACCTAAATCAACCATCTGCACCCCATCCAAACACAAACAGNATCAAAAGACACA	1972
Qy	268	AspSerLeuLeuLysGlnLeuTyrLeuProLeuSerGlnArgHisValGluSerAsp	287



JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

20499374  
11042159  
3

AUTHORS  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,  
Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kaehiwagi,K.,  
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,  
Yoneda,Y., Ishikawa,T., Okawa,K., Tanaka,T., Matsura,S., Kawai,J.,  
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer

TITLE  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

20530913  
11076961  
4

AUTHORS  
The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection

TITLE  
Nature 409, 685-690 (2001)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

5

AUTHORS  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

TITLE  
Nature 420, 563-573 (2002)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE

6 (bases 1 to 4561)

AUTHORS  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216]  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. 4561  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM,DB:E230012G01"  
/db\_xref="MG1:2425940"  
/db\_xref="taxon:10090"  
/clone="E230012G01"  
/sex="female"  
/tissue\_type="oviduct"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="2 days pregnant adult"

misc\_feature  
337..4560  
/note="laminin, alpha 4 (MGD|MG1:109321, GB|Y09827,  
evidence: BLASTN, 99%, match=4527)  
putative"

## ORIGIN

Alignment Scores:  
Pred. No.: 15.2 Length: 4561  
Score: 135.00 Matches: 140  
Percent Similarity: 34.91% Conservative: 104  
Best Local Similarity: 20.03% Mismatches: 257  
Query Match: 3.24% Indels: 198  
DB: 11 Gaps: 34

US-09-765-271-56 (1-796) x AK054014 (1-4561)

QY 8 GlnAlaArgThrValLysGlu-----AsnAsnArgValSerTyrIleAsp 22  
Db 1474 CAAGCCACACATGAGGACAAATAATCAAGAGATCAACAGCATGCTCTACTATGGA 1533

QY 23 GlyLysGlnAlaThrGlnLysThrProAspGluValSerLysArgGlu 42  
Db 1534 -----GAGAATCAGGAACCTTGGCCCGAGGAATAGCTGAGAAGCTG 1575

QY 43 GlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHis 62  
Db 1576 ---GTGTTGGCCCAAAAAATGCTGGAAGAGGTCCGAAGCCGAGCCATTCCTACCCAT 1632

QY 63 GlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGlu 82  
Db 1633 CGGAG-----CTGTTGGATGAGGAGGAGGATGAGGCCGAGGAA 1671

QY 83 LeuLeuMetLysAspProAsnTyrLysLeuLysAspGluLeuValValAsnGluValLys 102  
Db 1672 TTGCTGAGCAAGCTGAGAACTGG-----CAGCGCTGCACAAATGACACCCGC 1719

QY 103 GlyGlyTyrValIleLysValAsp-----GlyLysTyrTyrValTyrValLysAspAla 120  
Db 1720 TCITTTATCCCGTGGTCTGGAGCAGCTGACGACTACACGTAAGCTGTCCGACCTC 1779

QY 121 AlaHisAla-----AspAsnValArgThrLysGluGluIleAsnArg--- 134  
Db 1780 CAGGAATCCATTAAACAGGCCCTCGACCATGTCTAGGATGTCAGAAAGACATGAATAGAGCC 1839

QY 135 -----GlnLysGlnGluHisSerGlnHisArgGluGlyGlyThrProArgAsn 150  
Db 1840 ATCACTTTCAAGCAGCGGACCATGAGAAACAACATGAGAGTGAAGGAACACATGGAA 1899

QY 151 AspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIle 170  
Db 1900 GTTGTGGTGCGCTCTCTGAGTATGCTCTGACAGACTCTCTTACCATACCTCAGCTCACTCTT 1959

QY 171 PheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAsp 190  
Db 1960 GAGGAACCTTGATGAGATTAAGAATGTCATCTGGAATTTATGCAGAAATAGATGGA--- 2016

QY 191 HisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAla 210  
Db 2017 -----GCCAAAAATGAGCTCAAGGAAA----- 2040

QY 211 PheLeuSerGlyArgGlyAsnLeuSerAsn-----SerArgThr 223  
Db 2041 ---CTATCAACCTGAGTAACCTCAGTCATGCTTGGTTTCAGGAAGCTACGGACCATGCA 2097

QY 224 TyrArg---ArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSerValSer 242  
Db 2098 TACAATCTTCACAGGAGCCGATGAGCTAAGCAGAAATTTGCACATTCAGACATGAAC 2157

QY 243 AsnProGlyThrThrAsnThrAsnThrSerAsn---AsnSerAsnThrAsnSerGlnAla 261  
Db 2158 GGGCTGTCAGAGAAGGCTTTGGATGTCATCAACGCTCTATGAAATATCGCCAAATATGTC 2217

QY 262 SerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGln 281  
Db 2218 AGTGAAGCCCAACGAAACAGCAGAACTT----- 2244

QY 282 ArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAla 301



```

Db      2245  -----GCTCTGAATATCACTGATCAATTTAT 2271
QY      302  ArgGlyValAlaValPro---HisGlyAspHis-TyrHisPheIleProTyrSerGlnMe 320
Db      2272  GATGCTGTGAGTGGGAGTTGACACGACAGATCAATTTACCAT 2311
QY      320  tSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTr 340
Db      2312  -----AAGGATGAAGTGCACACCT 2331
QY      340  pValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGl 360
Db      2332  TCTCAATCAGCCAGCAGACTCGAGCCACAGCAGAT 2368
QY      360  yProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVa 380
Db      2369  -----TCTAGCAATGATGAAGCAGTGGCTGCACACCAG 2400
QY      380  lArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAl 400
Db      2401  CAGGCGTGTGGT---GGAGCCCTGTGGAGGAGGCGCCCTC 2440
QY      400  aLysAspLeuProSerGluThrValLysAsnLeu 411
Db      2441  -AGACACAGACTGAATGATGCTGTTAAGCAACTACAGGCAGCAGAGAGGGAGCGCCCA 2499
QY      412  -----GluSerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAl 427
Db      2500  CCAGCGCTCGGCGCAGTCCAGCTCTTCATTGAGGAAGCT---AACAGACACAGCAGCGC 2556
QY      427  aLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLe 447
Db      2557  TGTCCAAAGAGTTACACACCAATGCTAACACCTCACAACCTGCTCCCAACCTGCA 2616
QY      447  uThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLe 467
Db      2617  GACC-----TTTGACTCATCTGCATA 2637
QY      467  uAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAs 487
Db      2638  TAACACTGCGAGTGGACTGCTCGGGAGCGCAGTGAAGAACCTCACCAGGTTGTCCCCCA 2697
QY      487  pLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGl 507
Db      2698  GCTTCTGGATCATCTGCTGCTGAGCAGAG-----CGGCTCGACAGCAA 2745
QY      507  nIleGluTyrThrGluAspGluValArg-----IleAlaGln-----LeuAlaAs 522
Db      2746  CATTTCTGCAGCATCCAGAGGATCCGAGAGCTCAITGCTCAAACCCAGGAGTGTGCAAG 2805
QY      522  pLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGl 542
Db      2806  CAAGATCCCAAGTCTCC-----ATGATGTTTGTAT-----GG 2835
QY      542  vAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSe 562
Db      2836  CCAGTCAGTGTGCGAAGTGCAC-----CCCAAGTCACTGTGGA 2874
QY      562  rAsp-----LysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProPr 581
Db      2875  TGACCTGAAGGCGCTTCACATCCATCATGCTGTGTATCAAGACCT-----CCTCC 2922
QY      581  oSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAs 601
Db      2923  AAGGCG-----CGAGGCCACTGGGCGCTGGGTAGCAGATCAGTTTGT 2967
QY      601  nArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHisThr 621
Db      2968  CCTTACTCTCGAAGCAAAAC-----GCCAAAAGAAATACATG-----GG 3009
QY      621  rValGluValLysAsnGlyAsnLeuIleProHisLysAspHisThrHisAsnIle-L 641

```

---

```

Db      3010  TCTGCAATCAAAATGATGATACCTGGTATACCTTTTACATTTGGGGATGAAAGATGTGGA 3069
QY      641  yPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeu 658
Db      3070  AATTCTCTCTGGATTCCA-----AGCCTGTGAGCTCTCTGGCCGCTTACTTTA 3116

```

RESULT 20  
 CG753778/c  
 LOCUS CG753778  
 DEFINITION P049-1-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,  
 genomic survey sequence.  
 ACCESSION CG753778  
 VERSION CG753778.1 GI:37978601  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 1643)  
 AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,  
 Buntjer,J., van der Meulen,M. and Sommer,R.J.  
 TITLE An integrated physical and genetic map of the nematode Pristionchus  
 pacificus  
 JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)  
 MEDLINE 22835951  
 PUBMED 12884007  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1..1643  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Ppa EcoRI BAC Library"  
 /note="The library was generated by a partial digest of  
 the genomic DNA with EcoRI and cloning into the BAC  
 vector."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.56 Length: 1643  
 Score: 134.50 Matches: 119  
 Percent Similarity: 35.76% Conservative: 87  
 Best Local Similarity: 20.66% Mismatches: 219  
 Query Match: 3.23% Indels: 155  
 DB: 29 Gaps: 26

US-09-765-271-56 (1-796) x CG753778 (1-1643)  
 QY 191 HistTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAla 210  
 Db 1568 CACTACAACACTACTCACCACAA---CAATTATCATCAACACACAAATCAACACAG----- 1518  
 QY 211 PheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSerAsp 230  
 Db 1517 ---TAGCCACCACAAATCAGTCGCTAAGCAACACACATATCAACACAAACACGCAATA 1461  
 QY 231 AsnThr---SerArgThrAsnTrpValProSerValSerAsnPro----- 244  
 Db 1460 CAGACATAAACACACCCACAC---AAAAATATAAACACCAACCAACAGAAATACACN 1407  
 QY 245 -----GlyThrThrAsnThrAsn-----Thr 251  
 Db 1406 AAAAAACACACACCTCAATCAACAAATTTAAAGACCCGCGATAAACAATCCACAATAAAGT 1347  
 QY 252 SerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeu 271

```

1346 ACAACAAACAAATAATCNCACCTCAACCCCTCAGAGCATAGATACATATTT 1288
272 LysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPhe 291
1287 CAGAAATAGAACAAACCAACGAAAC--CAATACAGGCACAA-GAAGAAATAGGAAAGTCATA 1232
292 AspProAlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAsp-Hi 311
1231 GAACCG-ACCAACGAGACC-ACGCCAACAAACCAACAA 1196
311 sTyrHisPheIleProTyrSerGlnMetSerGluLeuGluGluArgIleAlaArgIleI 331
1195 CTACCCCAAGCACAAGCTCAAGACAGAGTCACACCACTTACCACACCCCAACAAACATA 1136
331 eProLeuArgTyrArgSerHisHisTyrValProAspSerArgProGluGlnProSerPr 351
1135 AAAAAAAAAAAAAAGCAAAACCAAAATATGCGAAGACATACATAAAAC-CCAAAC-- 1079
351 oGlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSe 371
1078 -----ACACCAAGCGCTTACCACCAACCAACCAACCAACCAACCAACCAACATA 1031
371 rAsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluG 391
1030 -AACACAAGCGCATAGAAACAA----- 1010
391 uLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLe 411
1010 ----- 1010
411 uGluSerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAs 431
1009 -AAAAAAAAAAAAAAAAAAAAACAAACACACACCCACCAACCAATATAAAAGCAACCC 951
431 nValAla-----ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuTh 448
950 AACCGCACACAGAAACCAACAAACAAACCAACCAACCAACCAACCAACCAACCA 891
448 rGluAlaHisLys-AlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeu 468
890 CCACACCCCAACACGCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 831
468 spLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspL 488
830 GGAACACAAACACACACGACGACCAACCAACCAACCAACCAACCAACCAACCA 790
488 euLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsnSerGlnI 508
789 -----AATCCCAACGACCAAAATTACAGCCCAACCAACCAACCA 756
508 leGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerA 528
755 GAAACAAACAGAGAGACAAACA-----CAAGAACCAAAATATATAACCAACACAG 706
528 spGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrP 548
705 AAGGAGAGAACTTCAAGACCAAC-----ACAA 679
548 roHisMetGlyHisSerHisTyrIleGlyLysAspSerLeuSerAspLysGluLysVala 568
678 GACACAAACCCACACAAACAC--ACAACAAATTTACCATACCAACCAACCAACTCAAAACAA 622
568 laAlaGlnAlaTyrThrLysGluLys-GlyIleLeuProProSerProAspAlaAspVal 587
621 ACACCAAAATGTATCATGAAGCAAAATAAATCTATCGAAAAAATACACCTCAAAACAAATA 562
588 LysAlaAsnProThrGlyAspSerAlaAlaIleTyr----- 600
561 CCACCAACCAAGCCAGACCCCAAGCATCTCAACAACTACCAACCAACCAACCAAGAAA 502
601 ---AsnArgValLysGlyGluLysArgIleProLeuVal-----ArgLeuProTyr 616

```

```

501 CACCACCGCTAATAAAAAAAAAAAAAAAAAAACTACCCATGTATCTCTCATATGTCCCCAC 442
617 Met-----ValGluHisThrVal---GluValLysAsnGlyAsnLeuIlePro 632
441 CAACAACCCCAATACACACCAATAATAGAAATGTTTAAATCTCAAGATACCAATACCA 382
633 HisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAla 652
381 CACAATGAC-----CAACATACAA-----AAACACACACACCCCTCTG 343
653 ProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisPro 672
342 CCCCCCACCACACATATAAAAAA-----AAACAACCG 310
673 AspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys 692
309 GACAAA-----ACACAACACCCACACACAAA 283
693 LysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGluProValGluGlu 712
282 CAACAACATCAACATCCCCATCAACAAACGAAAGTACCAAGAAATACACCCACACACA 223
713 ThrProAlaGluProGluValProGlnValGluThrGluLys 726
222 CAACCC---AAACCAAAACAAACAAACCAACCAACCAACAA 184

RESULT 21
CG756662/c 1652 bp DNA linear GSS 24-OCT-2003
LOCUS P051-4-D06.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG756662
VERSION CG756662.1 GI:37984450
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1652)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES
source 1..1652
Location/Qualifiers
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Alignment Scores: 3.59 Length: 1652
Pred. No.: 134.50 Matches: 116
Score: 30.17% Conservative: 65
Percent Similarity: 19.33% Mismatches: 218
Best Local Similarity: 3.23% Indels: 203
Query Match: 29 Gaps: 24
DB:

```



Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
Class: BAC ends.

## FEATURES

```

1. 1795
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by
the genomic DNA with EcoRI and clon

```

## ORIGIN

Alignment Scores:		
Pred. No.:	3.94	length:
Score:	134.50	Matches:
Percent Similarity:	33.46%	Conservative:
Best Local Similarity:	20.46%	Mismatches:
Query Match:	3.23%	Indels:
DB:	29	Gaps:
		19
		1755

US-09-765-271-56 (1-796) x CG754413 (1-1755)

219	QY	SerAsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpVal	238
1601	Db	AAACAACAAATAGAAATATATACACAACAACAAAAAGACAACAACCACTAATCAAGTA	1542
239	QY	ProSerValSer---AsnProGlyThrThrAsnThrAsnThrSerAsnSerAsnThr	257
1541	Db	ACTTNCCTGAAATAAATAATCCAAACAT-ACAACAACAACAACACCAACCAACAGGATACT	1483
258	QY	AsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln-----	273
1482	Db	AAATCATCCCCAAACACACACACACATTCCTAACCAATCATCAC-AAACAATATCACCAGCAAT	1424
274	QY	-----LeuTyrLysLeuProLeuSerGlnArgHisValGluSer	286
1423	Db	AAAAACACACAGTACACAACAATATATACAGAAAAACCCAGACAAGACACACAC-ACAAC	1365
287	QY	AspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaargGlyValAlaVal	306
1364	Db	GAT-----CCACCACACCCACACAACACCCAGAACCCCAACGATATACAACACC	1317
307	QY	ProHis-----GlyAspHisTyrHisPheIleProTyrSerGlnMetSerGlu	322
1316	Db	CACCACCCTCCCCACACAACAAGCAACTACCGCACACACACATAAACAACACCAAAA	1257
323	QY	LeuGluGluArgIleAlaargIleIleProLeuArgTyrArgSerAsnHisTrpValPro	342
1256	Db	CTAAACAACAAC-ACA-----AACTCAAGCGCAAAACACAA	1222
343	QY	AspSerArgProGluGlnProSerProGlnProThrProGluProSerProGlyProGln	362
1221	Db	CAACAAACCAACAACAACAACAACAACAACAACAATACAGCCCATTA-CAACACAA	1163
363	QY	ProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLys	382
1162	Db	CCAAAGACACATCA-CACATCAACAGCACCATTAGACACAACCATCAATAAAACGAGCA	1104
383	QY	ValGlyGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAsp	402
1103	Db	CAGCGGACATTATACAACACGAAAGTNGCAATCATCAAAATATATAGTACACAACCA	1044
403	QY	LeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSer	422
1043	Db	CACACCAACCAACACC-----TCCAAATATCCAGCAACAACCCCAACATC	996
423	QY	HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLys	442

995	Db	CAC	TAT	AAAGAA	AAACCAAGCAAAATCA	-----AGT	966									
443	QY	Ala	Tyr	Asn	Leu	Thr	Gl	uAla-His	Leu	Phe**Asn	Lys	Gly	Arg	Asn	Se	462
		:::														
965	Db	TC	AT	CA	AA	CT	T	ACT	CA	T	CA	T	CA	T	CA	908
462	QY	rAsp	Phe	Glu	Ala	Leu	Asp	Lys	Leu	Glu	Arg	Leu	Asn	Asp	Glu	482
		:::														
907	Db	---	---	---	---	---	---	---	---	---	---	---	---	---	---	885
482	QY	uLys	Leu	Val	Asp	Asp	Leu	Leu	Ala	Phe	Leu	Ala	Pro	Leu	Thr	497
		:::														
884	Db	CC	AA	CT	CC	CA	CC	CA	CT	CC	TC	AC	CC	CT	TC	825
		:::														
498	QY	Pro	Glu	Arg	Leu	Gly	Lys	Pro	Asn	Ser	Gln	Ile	Glu	Tyr	Thr	514
		:::														
824	Db	AC	AA	CC	CA	AA	CT	CA	CA	CT	CA	CA	CT	CG	AA	765
		:::														
515	QY	Val	Arg	Ile	Ala	Gln	Leu	Ala	Asp	Lys	Tyr	Thr	Ser	Asp	Gly	533
		:::														
764	Db	CAC	CA	CT	CA	AA	CT	CC	CT	CA	CA	CG	CA	CA	CA	705
		:::														
533	QY	pGlu	His	Asp	Ile	Ser	Asp	Glu	Gly	Asp	Ala	Tyr	Val	Thr	Pro	553
		:::														
704	Db	AA	CA	CA	CA	CA	AA	TA	-----AA	AG	CA	CA	CC	AT	CA	666
		:::														
553	QY	rHis	T	Pro	Ile	Gly	Lys	Asp	Ser	Leu	Ser	Asp	Lys	Glu	Lys	573
		:::														
665	Db	TC	CA	AA	CT	CA	-----GC	AA	GT	TAC	AC	CA	CA	CA	CA	642
		:::														
573	QY	rLys	Glu	Lys	Gly	Ile	Leu	Pro	Pro	Ser	Pro	Asp	Ala	Asp	Val	593
		:::														
641	Db	TACT	CA	CC	CG	CAG	CA	CA	AC	CC	CA	CT	CAC	CG	AAG	584
		:::														
593	QY	Y	Asp	Ser	Ala	Ala	Ala	Ile	Tyr	Asn	Arg	Val	Lys	Gly	Lys	613
		:::														
583	Db	-----CAC	AT	AA	T	CT	CG	AT	CT	CT	CT	CT	CT	CT	CT	566
		:::														
613	QY	gLeu	Pro	Tyr	Met	Val	Glu	His	Thr	Val	Glu	Val	Lys	Asn	Gly	633
		:::														

RESULT 23

RESUL 23  
AY411030

HY411030  
LOCUS

## FOCUS DEFINITION

## DEFINITION

ACCESSION

ACCESSION  
VERSION

4773 bp DNA linear GSS 16-DEC-2003  
 , VIRTUAL TRANSCRIPT, partial sequence,

KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 4773)  
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 4773)  
AUTHORS Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.  
FEATURES  
source  
1..4773  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>4773  
/gene="RBBP6"  
/locus\_tag="HCM4088"  
ORIGIN  
Alignment Scores:  
Pred. No.: 17.9 Length: 4773  
Score: 134.50 Matches: 137  
Percent Similarity: 34.10% Conservative: 131  
Best Local Similarity: 17.43% Mismatches: 304  
Query Match: 3.23% Indels: 214  
DB: 29 Gaps: 28  
US-09-765-271-56 (1-796) x AY411030 (1-4773)  
QY 24 LysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGluGly 43  
DB 2581 AAAGAGGAGCTCCGAGGACTGCAATACATACTAATCATCTCTCTCAGAGGAGTAA 2640  
QY 44 IleAsnAlaGluGlnIleValIleLysIleThrAsp-----GlnGlyTyrValThr 60  
DB 2641 -----AAATCACTGGAAACCCCGAGAAAGCTCACTCTAAA 2676  
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
DB 2677 TCAGCAAAAGACACCAAGAA-----ACAAACCCAGTCAAA 2712  
QY 81 GluGluLeuLeuMetLysProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
DB 2713 GAGGAAAGAGTGAAGAGGACTATTCGAAAGTGTCAATCAGAAAGCTCACTAAG 2772  
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAspAla 120  
DB 2773 -----GAAGAAAG 2781  
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
DB 2782 GCCAAGAACCTAATCAGAAACCAACCACTTGATAATAGGAGGAGGAGGAGGAGGAGG 2841  
QY 141 GlnHisArgGlyGlyGlyThrProArgAsnAspGlyValAlaLeuAlaArgSerGln 160  
DB 2842 AAAACTGAAGAAAGGCGTAGATAAGATTTTGAGCTCTCTCAATGAAATCTCGAA 2901  
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180

2902 -----CTAGAAGTGACTGAAATAGTGAACCATCA 2931  
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIlePro----LysAsnGlu 199  
DB 2932 CCAAAAGCGCAAAATGGAACCTGATCTGAAAAAATGGATAGACCCCTGAAAGGACAAA 2991  
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
DB 2992 ATTTCCTTAACT-----GCGCAGCCCAAAAAATCAACTCAACAGA----- 3033  
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
DB 3033 ----- 3033  
QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
DB 3034 -----GAAACTGGGGAAGAAATTTGGAAGTACAGAAAAATATATCAACACCAAGAA 3084  
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln---LeuTyrLysLeuPro 278  
DB 3085 CCTCTGAAAAATTTGGAGTCAACATCTAGCAAGTTAAACAGAAAAAGTCAAGGAAAG 3144  
QY 279 LeuSerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSer 298  
DB 3145 GTCAGACGAAAGTCTGCTGGAAGTCAAGGATCCAGCTCAACTCTGTGTGATTACACAGT 3204  
QY 299 ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSer 318  
DB 3205 ACGAGCTCAACTGGAGGCGAGT-----CCTGCGCG 3234  
QY 319 GlnMetSerGluLeuGluGluAlaArgIleAlaArgIleProLeuArgTyrArgSerAsn 338  
DB 3235 AAATCTGAAGAAACACAGATACAAAGCCAACTGTGATTAAACAGTGAAGAAATATA 3294  
QY 339 His-----TrpValProAspSerArg 345  
DB 3295 AATGACAAATACCGCGCGAGCTGAAGATGTTATCATTTATGATTCAGGTTCTCAATCCAAA 3354  
QY 346 ProGluGlnProSerProGlnProThrProGluProSerProGlyProGlnProAlaPro 365  
DB 3355 TGGGATAAAGATGACTTTTGAATCTGAAGAAAGATGTTAAATCCACACGCTATATCA 3414  
QY 366 AsnLeuLysIleAspSerAsnSerLeuValSerGlnLeuValArgLysValGlyGlu 385  
DB 3415 AGT-----GTAGGAAACCTCTAGTGTATATAAAATGTTAGTACAAAG----- 3459  
QY 386 GlyTyrValPheGluGluLysGlyLysSerArgTyrValPheAlaLysAspLeuProSer 405  
DB 3460 -----CCATCAAAATATAGTCAAGTATCTCTGAGAAAGAAAGTGAAGCATCC 3504  
QY 406 GluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHisThrLeu 425  
DB 3505 GAGAAAAATTCAGAAATTCACCAAGGACGTGAGC-----CATGAATCATCAACATGAGGT 3561  
QY 426 ThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsn 445  
DB 3562 AAAAGTTCAAAAAATCTCTCATCTAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 3621  
QY 446 LeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGln 465  
DB 3622 GTGTGGAAAAAGAGAACCT-----GAAAAGAGGAGGAGGAGGAGGAGGAGGAGG 3669  
QY 466 AlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLysLeuVal 485  
DB 3670 GAGAAAGAGAGTAAATTTGACCGCTGGAATGACACAGGAAATTTTAAAGT----- 3720  
QY 486 AspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLysProAsn 505  
DB 3720 ----- 3720  
QY 506 SerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThr 525

Db 3721 -----CTGTCTCAATCTTCCAAAGAGGCTAGA 3747

Qy 526 ThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyr 545  
 |||||  
 : : : : :  
 Db 3748 ACGTCA-----GATAAATCATGATCCACTCGTCTCTCTCAATAAAGAC 3792

Qy 546 ValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSerLeuSerAspLysGlu 565  
 |||||  
 : : : : :  
 Db 3793 TTCATCCCAATAGA-----GACAAAACCTGACTATGACACACAGA 3834

Qy 566 LysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProProSerProAspAla 585  
 : : : : :  
 : : : : :  
 Db 3835 GAGTATTCAGTTCCAAACGCTAGACATGAAAAGATGAATTAACAAGACGAAAGACCTCT 3894

Qy 586 AspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGly 605  
 |||||  
 : : : : :  
 Db 3895 CCTTCTCGG-----ATAAAGATTTCTGATCTGGACAGAAAATAAACCAAGGAA 3945

Qy 606 GluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLys 625  
 |||||  
 : : : : :  
 Db 3946 GAGAGAGATTTGCCT-----AAAAAGGACAGGAGATTCACAA 3984

Qy 626 AsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPhe 645  
 |||||  
 : : : : :  
 Db 3985 AAAAGTAAT--TCTAGTCCCTCAAGACAGACA----- 4014

Qy 646 AspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIle 665  
 |||||  
 : : : : :  
 Db 4015 AAACCTCATGATCAAAAGCC----- 4035

Qy 666 LysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp-----GlyTyrGly 683  
 |||||  
 : : : : :  
 Db 4036 ACTTATGATATAACGCCAAATGAAGACAGACAAATCTGTAGATAAAAATCCTTGTAA 4095

Qy 684 AsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsnPhe 703  
 |||||  
 : : : : :  
 Db 4096 GATCGTGAGACGATGTATTAGAAGCAAGGAACAATAAAGTCAAGTGGCAATAAAGCTA 4155

Qy 704 LysAlaAspGluGluProValGluGluThrProAlaGluProGluValProGluValGlu 723  
 |||||  
 : : : : :  
 Db 4156 CTTTATATATCTTAAACCCACCCAGAG-----ACACAGGTTGAA 4191

Qy 724 ThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeu--LeuAlaLysValThr 742  
 |||||  
 : : : : :  
 Db 4192 AAAGAGCAATTAATGCGCAATTCACAGAGTACTGTCAAGCTTAACCCAGTTAGT 4251

Qy 743 AspSerSer---LeuLysAlaAsnAlaThr--GluThrLeuAlaGlyLeuArgAsnAsnLe 761  
 |||||  
 : : : : :  
 Db 4252 CATTCCTCTAGACTTCTCTGACTTAAGTAACTGAGAAACTG----- 4291

Qy 761 uThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLe 781  
 |||||  
 : : : : :  
 Db 4292 ---ATGAAGCTGCTTTTGAACCAAGACTATAATGAAAGTGACAGTGAAGTAATGTTTCT 4347

Qy 781 uLeuLysGlySerAsn 786  
 |||||  
 : : : : :  
 Db 4348 GTAAAGAGAGAGAT 4363

## RESULT 24

CG755974/c

LOCUS

DEFINITION P051-2-E12.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CG755974

VERSION CG755974.1

KEYWORDS GI:37983025

SOURCE GSS.

ORGANISM Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Nemodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 2041)

AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.

TITLE An integrated physical and genetic map of the nematode Pristionchus

JOURNAL pacificus

MEDLINE Mol. Genet. Genomics 269 (5), 715-722 (2003)

COMMENT 22835951

PUBMED 12884007

CONTACT: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

Class: BAC ends.

location/Qualifiers

1. 2041

/organism="Pristionchus pacificus"

/mol\_type="genomic DNA"

/strain="Californica"

/db\_xref="taxon:54126"

/clone\_lib="ppa EcoRI BAC library"

/note="The library was generated by a partial digest of

the genomic DNA with EcoRI and cloning into the BAC

vector."

## ORIGIN

Alignment Scores:

Pred. No.: 5.43 Length: 2041

Score: 134.00 Matches: 122

Percent Similarity: 29.39% Conservative: 77

Best Local Similarity: 18.02% Mismatches: 256

Query Match: 3.22% Indels: 223

DB: 29 Gaps: 25

US-09-765-271-56 (1-796) x CG755974 (1-2041)

Qy 172 AsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIle-----ValProHisGly 189

Db 2002 ANTACAAACACTACACACACACACACATCGCTACATACACACAGACCGCACAA 1943

Qy 190 AspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGlu 209

Db 1942 AGATCGCAACACACACACACCT----- 1922

Qy 210 AlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgGlnAsnSer 229

Db 1921 ----- 1913

Qy 230 AspAsnThrSerArgThrAsnTyrValProSerValSerAsnProGlyThrThr----- 247

Db 1912 ACACACAAACCAACACACACACACCGCCACATCGCAACACACACACCTTGTAA 1853

Qy 248 AsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIle 267

Db 1852 CACACACACACCAACACCGACACACACACACACCGCCACACACACACACACACATA 1793

Qy 268 AspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAsp 287

Db 1792 CACGCA-----TCCTCGCACACACACACACGACG--- 1766

Qy 288 GlyLeuValPheAspProAlaGlnIleThrSer-ArgThrAlaArgGlyValAlaValPr 307

Db 1765 -----CCACCCACACTACACACACCGCCACACACACACACACACCCAC 1721

Qy 307 oHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGlnLeuGluArgIle 327

Db 1720 ACACACAGACACCAAAATACCGCCCGCCACCTCCACAAACACCA----- 1675

Qy 327 eAlaArgIleIleProLeuArgTyrArgSerAsnHisTyrValProAspSerArgProGl 347

Db 1674 -----CCACACACACCAACCGACCGAC---AC 1649

Qy 347 uGlnProSerProGlnProThrProGluProSerProGly----- 360

Dd	1648	ACAACCAACACACACCAACACACACACACACCAACCCACCTCTACCCCGCCGCAC	1589
Qy	361	----ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLe	379
Dd	1588	ATAGACCAACCAACCAACGAC	1567
Qy	379	uValArgLysValGlyGluGlyTyrValPheGluGluGlyGlyIleSerArgTyrValPh	399
Dd	1566	-----ACACACCAACACAAACAAACACTCTCTACAGCGGTGCA	1532
Qy	399	eAlaLysAspLeuProSerGluThrValLysAsnLeuGlySerLysLeuSerLys	417
Dd	1531	CAGCCCCACATGCTCCCAACATAAABAAAC--AAABAAAACACAAAATAAA	1475
Qy	418	-----GlnGluSerValSerHisThrLeuThrAla	427
Dd	1474	AACATAACAGACCATCTATCCCAACACACAAACCGCCCCAACGCCACCCACCCGCGCAC	1415
Qy	427	-----	427
Dd	1414	ACCATAGACACCGCAGACATCTACGACACACCCACCCANAGCCGACCTCTACNACCACA	1355
Qy	428	-----LysLysGluAsnVal	434
Dd	1354	AACCAACCAACAAACAAAAACATAAAAAATACATCAAGCCACACAGAACCCACAC	1295
Qy	434	oArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLe	454
Dd	1294	TCACACAAAGCGAACTATCGCGNCGCG--ACCCAGACAGCAACAAAT	1247
Qy	454	uPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLe	474
Dd	1246	ATACAAACCAACAAACACACAAACACACACACGCAACATGCAACCAACAAAAACACACAA	1187
Qy	474	uAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPr	494
Dd	1186	CCAGACCCACAAACACAGACCAACCAACCAACCAACANAGACCACACA	1136
Qy	494	oIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspG	514
Dd	1135	AACAAACACAAAGACACCACTAACACAGACACACACACACCAACCAACCAACTAGCA	1076
Qy	514	uValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAsp	532
Dd	1075	CACACATATCAACACTCTACACGCGACACACACACACACAAATACAGCGCACACACACA	1016
Qy	532	eAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHi	552
Dd	1015	CAAAACACACACACACAGCACACACACCCCAACACACACACACACACACACACAGGA	956
Qy	552	sSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTy	572
Dd	955	ACAGCATACGTACTCTACCACCACTCTCCAGAAC	922
Qy	572	rThrLysGluLysGlyLleLeuProProSerProAspAlaAspValLysAlaAsnProTh	592
Dd	921	-----CCAACAAACGAAACCAACTCATTAACAAACACACAAAC	887
Qy	592	rGlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuVa	612
Dd	886	ACA--CAACACAGTCCACGCGCCACCAACCATACAG	846
Qy	612	lArgLeuProTyrMetValGluHiSThrValGluValLysAsnGlyAsnLeuIleLePr	632
Dd	845	AAGAAACGCAAAATCTACACACACACAAACACCCAGCCACCACTCATCAACACACAC	786
Qy	632	oHisLysAspHisThrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLys--	651
Dd	785	AAGAACGCAATATAAACCAACAAACAA	732
Qy	652	-----AlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr	664
Dd	731	CAGCACCAACACAGAGCGGGGAGACATAAACACACACACCAACACGCAACAAAAACGA	672

Qy	665	-----IleLysTyrTyrValGluHisProAs	673
		: : : :	
Db	671	CACAAAGACCAACCCACACAAAAAATCATCCACACACACCA	612
Qy	673	pGluArgProHisSerAsnAspGlyTrrpGlyAsnAlaSerGluHisValLeuGlyLysLy	693
		: : : :	
Db	611	CCACGACCAACACCCCAAAACACAATAAGAACAACACACAC	566
Qy	693	sAspHisSerGluAspPro	699
		: : : :	
Db	565	-CCGCACACCAACAAACCCCCCTGACCTCCACAAACCAACATATACATGAAAG	507
Qy	700	-AsnLysAsnPhelysAlaAspGluGluProValGluGluThr	713
		: : : :	
Db	506	AACGAATAACAAACAACCGGTACCGAACCAACTCACACAAACACCAACCAATCCA	447
Qy	714	-ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysG	733
		: : : :	
Db	446	ACCACCAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAGGA	387
Qy	733	uAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsn	749
		: : : : : : :	
Db	386	GAGACAA-----CTACAAAAAACACACACCAACCAACCAACCAAAATCTCAT	344
RESULT 25			
AY404466			
LOCUS			
DEFINITION Homo sapiens HCM1900 gene, VIRTUAL TRANSCRIPT, partial sequence,			
AY404466			
3714 bp DNA linear GSS 12-DEC-2003			

RESULT 25	AY404466	3714 bp	DNA	linear	GSS 12-DEC-2003
LOCUS	AY404466				
DEFINITION	Homo sapiens HCM1900 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.				
ACCESSION	AY404466				
VERSION	AY404466.1	GI:39760443			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1. (bases 1 to 3714) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science	302	(5652),	1960-1963	(2003)
PUBMED	14671302				
REFERENCE	2. (bases 1 to 3714) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sminsky,J.J., Adams,M.D. and Cargill,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.				
FEATURES	Location/Qualifiers				
source	1..3714 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" <1..3714 /locus_tag="HCM1900"				
gene					
ORIGIN					
Alignment Scores:					
Pred. No.:	13.4	Length:	3714		
Score:	134.00	Matches:	183		
Percent Similarity:	34.60%	Conservative:	135		
Best Local Similarity:	19.91%	Mismatches:	350		
Query Match:	3.22%	Indels:	254		
DP:	29	Gaps:	41		

Alignment Scores:		
Fred. No.:	13.4	Length: 3714
Score:	134.00	Matches: 183
Percent Similarity:	34.60%	Conservative: 135
Best Local Similarity:	19.91%	Mismatches: 350
Query Match:	3.22%	Indels: 254
DR:	29	Gaps: 41



US-09-765-271-56 (1-796) x AY404466 (1-3714)

QY 31 GluAsnLeuThrProAspGluValSerLysArgGluGlyIleAsnAlaGluGlnIleVal 50  
 Db 7 GAGAGGATACCAAGAGAGAGGTGCCCAAGAAAGATGGAGTT-----48  
 QY 51 IleLysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyrHisTyrTyrAsn 70  
 Db 49 -----GGTGATGCAACATTTTCGATTCC 72  
 QY 71 GlyLysValProTyrAsp---AlaIleIleSerGluGluLeuLeuMetLysAspProAsn 89  
 Db 73 AGCAAGATTATGGAAGCATAGTACATACAAAAGGAGCTAGATTTACAA-----AAT 126  
 QY 90 TyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysVal 109  
 Db 127 GGTAGCTTAGGAAGAGACTCTGTGGTGCAATCTGTGAGAACGATTCCTCCAAACATGATG 186  
 QY 110 AspGly-----LysTyrTyrValTyrLeuLysAspAlaAlaHisAlaAspAsn 125  
 Db 187 GAGAGCCTCAGCCCAAGAAATACTCTCCAGTCTGAGATTAAAGCCATGGAGACTAT 246  
 QY 126 ValArgThrLysGluGluIleAsnArgGln---LysGlnGluHisSerGlnHisArgGlu 144  
 Db 247 TCTGGCTCCTATTAAACCCCTCTCAACACCTGTGCTGCAAGAGAGCCCTTCTCCTTTG 306  
 QY 145 GlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGly---ArgTyr 163  
 Db 307 GGAACAGTGTCAAGAC-----AGCCCTCTCTAGCCAAATCCAGGGGAAGCAAGCAG 360  
 QY 164 ThrThrAspAspGlyTyrIlePheAsn-----AlaSerAspIleIleGlu 178  
 Db 361 TTCTCTTATGATGGAAGTGCACAAATAATCTCTATGAACTCCCACTCTTACTCAAC 420  
 QY 179 AspThrGly-----AspAlaTyrIleValProHisGlyAspHisTyrHisTyrIlePro 196  
 Db 421 ACTACATCTCCCTCAGTGGATATCCACTTGGAAAGAGCAGACTTTGATCATATACTGGC 480  
 QY 197 LysAsnGluLeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGly 216  
 Db 481 CGGGAC-----AGTGAAGGGCCCTG-----501  
 QY 217 AsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsn 236  
 Db 502 AGGCTCTCAGAGAGCCTCCCTATTCCAATATAGCTCAAGCCATAAATCGCATGACAA 561  
 QY 237 -----Tyr 237  
 Db 562 GTCTACTCTCTTGAGGGCTGGAAGTTCGGAAGGCATCTGGCTCGCTCCTGGCCATGTGG 621  
 QY 238 -----ValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253  
 Db 622 AATGGAAGTTCCCTTGATGTATGTGGCCGCTCTATATCAGCAGATCGGAGCGCGCAAGC 681  
 QY 254 AsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGln 273  
 Db 682 ATGCCTTCAGCCCAAGCAAGCAGCAAAATAGCATTAGCAGACAGCCTGGCGCTTCAA 741  
 QY 274 LeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspPro 293  
 Db 742 -----CCCAAGTTAACTAGACACAAAGGAG-----CTTGATCTGAAAC 780  
 QY 294 AlaGlnIleThrSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHis 313  
 Db 781 ATCAATTGAGAACTAGGAAGTACTCCAGCAGCAGCCTGAGTCACATGGGAGCCCTAC---837  
 QY 314 PheIleProTyrSerGlnMetSerGluLeuGluArgIleAlaArgIleIleProLeu 333  
 Db 838 -----AGCCCATCATTCCCAAG 855  
 QY 334 ArgTyrArgSer-----AsnHisTyrValProAspSerArgProGluGlnProSerPro 351  
 Db 856 TTGTACAGAGCCACAGAGAACCCAGCTGACACCTCTCAGCTTGCCTCCCAAGAAATCTCTCTG 915

QY 352 GlnProThrProGluProSerProGlyProGlnProAlaProAsnLeuLysIleAsp---370  
 Db 916 GGCATTTCCAAACGAAACAAACTTTGGGAAAGGATCTACCTCATAGCGTAATAGACAAT 975  
 QY 371 -----SerAsnSerSerLeuValSerGlnLeuValArg 381  
 Db 976 GACATTTACCTTAATTTTCTTTTCTTTGAGCTCAGGGGCTTTACCCCTATAAAACCTCTGCT 1035  
 QY 382 LysValGlyGluGlyTyrValPheGluGluLysGlyIle-----SerArgTyrValPhe 399  
 Db 1036 TCTGAAGGCAATCCTTAATGTAAGTTTACCTCAGTGTCCCTGCGCAGTCCAGGAGTGCT 1095  
 QY 400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
 Db 1096 CGGAAGATGCTTCTGGCTCCACCTCTCTGCTGCTCTGATGACTTTTGATCAGGCTTCA 1155  
 QY 420 SerVal-----SerHisThrLeuThr-----Ala 427  
 Db 1156 TATGTGGGACAAACCCGAGTCAATCTTCTGCTGGAGAGTCAGACAGAGTTTTCGG 1215  
 QY 428 LysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeu 447  
 Db 1216 ACCAGGAGGAATCTCTCTGTGATCTGTGGAATTTGATGAGGAGATTTTGGAAAGCCTC 1275  
 QY 448 ThrGluAla-----HisLysAlaLeuPhe\*\*\*AsnLysGlyArgAsnSerAsp 463  
 Db 1276 AGACGCTCAGGAACCCCGACCTGCGCTTTCGGGAACGGAAGACGATTTAGTCTCC 1335  
 QY 464 PheGlnAlaLeuAspLysLeuLeu-----GluArgLeuAsn 475  
 Db 1336 ATTTGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1395  
 QY 476 AspGluSerThrAsnLys-----GluLysLeuValAspLeuLeuAlaPheLeuAla 493  
 Db 1396 GAGCAGAAATGGAGCAGATTTGGAGAGACAGCGTCTGGAGACCATCTCAGTCTCTGTGCT 1455  
 QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGlu-----Tyr 510  
 Db 1456 GAATACACA-----AAGCCTGACAGTCTGCTTATCTACTCTGAGACCAACC 1497  
 QY 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530  
 Db 1498 GTGAAGATGTGAGAAATCAACAGGAGCTTGAGAGCTTGAGAGCTCTCTGATGAGAG 1557  
 QY 531 IlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMet 550  
 Db 1558 TCTGTGTTGAGGAAGCCCTCATGAGCCCTGAC-----ACAAGATACAGG 1602  
 QY 551 GlyHisSerHisTyrIleGlyLysAspSerLysSerAspLysGluLysValAlaAlaGln 570  
 Db 1603 TGCACCGG-----AAAGACTCCCTCCCTGATGATGATGATGATGATGATGATGATG 1650  
 QY 571 AlaTyrThrLysGluLysGly-----IleLeuProProSerProAspAlaAsp---Val 587  
 Db 1651 AGTCTCAGTCAGCAGTGCAGCTCTTTTACCCCGAGGAGCACCAGGATGATGAACTA 1710  
 QY 588 LysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGlyLys 607  
 Db 1711 CTCAGTGA-CCTCACCAGGACTCTCCACCATCTCTCCACCTTTCCGAAAGCTTCAG 1769  
 QY 608 ArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluValLysAsnGly 627  
 Db 1770 CGAGT-CCTCTTATTAAGTATCTCTACCAAGACCCAGAGGGGTATAAGTGAAGACAGA 1828  
 QY 628 AsnLeuIleIleProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAsp 647  
 Db 1829 GATCTC-----AGGAGTTGCTGCAAA---1849  
 QY 648 HisThrTyrLysAlaProAsnGlyTyr-----ThrLeuGluAspLeuPheAlaThr 664  
 Db 1850 -----TGGAAAGAACCCCGGATA-GTCATTTCTGAACAACTCTCGAGGAACTTAAGCAAAA 1902

Search completed: October 1, 2004, 11:30:12  
Job time : 5074 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 1, 2004, 07:39:37 ; Search time 139 Seconds

(without alignments)  
3177.992 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTVKNRVSY.....KLIALIKGNPSSVSKEKIN 796

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool.h/US09765271/runat 30092004 113754 12303/app query.fasta\_1.967  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=90 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=25  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09765271.qcgn1\_1\_69 @runat 30092004 113754 12303 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	2389	3	US-08-961-083-55
2	4163	100.0	2389	4	US-09-536-784-55
3	4163	100.0	2451	4	US-09-468-656A-9
4	4163	100.0	8195	4	US-08-961-527-94
5	2786.5	66.9	2531	4	US-09-468-656A-11
6	2772	66.6	2531	4	US-09-468-656A-5
7	2649.5	63.6	2290	3	US-08-961-083-65
8	2649.5	63.6	2290	4	US-08-961-083-65
9	1774	42.6	2359	4	US-08-961-527-243
10	1222.5	29.4	6867	4	US-08-961-527-192
11	1203.5	28.9	1455	4	US-09-468-656A-7
12	1187.5	28.5	1342	3	US-08-961-083-181

13	1187.5	28.5	1342	4	US-09-536-784-181	Sequence 181, Appl
14	988	23.7	973	4	US-08-961-527-355	Sequence 355, Appl
15	770.5	18.5	1684	4	US-08-961-527-258	Sequence 258, Appl
c 16	304.5	7.3	841	2	US-08-743-637B-34	Sequence 34, Appl
c 17	304.5	7.3	841	3	US-08-526-840B-34	Sequence 34, Appl
18	170	4.1	8155	4	US-08-956-171E-63	Sequence 63, Appl
19	169	4.1	12173	4	US-08-956-171E-310	Sequence 310, Appl
20	163	3.9	11091	4	US-09-134-001C-2243	Sequence 2243, Appl
21	159	3.8	30549	4	US-09-134-001C-322	Sequence 322, Appl
22	152.5	3.7	31096	4	US-08-956-171E-59	Sequence 59, Appl
23	150.5	3.6	2616	4	US-09-417-197-108	Sequence 108, Appl
24	148	3.6	14736	4	US-08-961-527-171	Sequence 171, Appl
25	147.5	3.5	2745	4	US-08-817-514F-5	Sequence 5, Appl
c 26	147.5	3.5	15213	4	US-08-961-527-26	Sequence 26, Appl
27	147	3.5	2439	4	US-09-513-783A-5	Sequence 5, Appl
28	146	3.5	3095	6	5231168-1	Patent No. 5231168
29	146	3.5	5400	4	US-09-134-000C-1773	Sequence 1773, Appl
30	145.5	3.5	4223	3	US-09-541-782-5	Sequence 5, Appl
31	145.5	3.5	4223	4	US-09-723-820-5	Sequence 5, Appl
32	145.5	3.5	4223	4	US-10-270-085-5	Sequence 5, Appl
33	144	3.5	3121	3	US-08-961-083-117	Sequence 117, Appl
34	144	3.5	3121	4	US-09-536-784-117	Sequence 117, Appl
c 35	144	3.5	14066	4	US-09-601-198-56	Sequence 56, Appl
36	143	3.4	4522	4	US-08-956-171E-39	Sequence 39, Appl
37	143	3.4	8100	4	US-09-554-337-4	Sequence 4, Appl
38	143	3.4	11517	1	US-07-920-281C-1	Sequence 1, Appl
39	143	3.4	11517	3	US-08-466-277-1	Sequence 1, Appl
40	143	3.4	15538	4	US-09-554-337-1	Sequence 1, Appl
41	141.5	3.4	3945	4	US-09-200-650B-6	Sequence 6, Appl
42	141.5	3.4	7941	4	US-09-816-703A-1	Sequence 1, Appl
43	139.5	3.3	4948	4	US-09-562-702A-23	Sequence 23, Appl
44	139.5	3.3	4948	4	US-09-561-818A-23	Sequence 23, Appl
45	139.5	3.3	4972	4	US-09-562-702A-27	Sequence 27, Appl
46	139.5	3.3	5306	4	US-09-562-702A-21	Sequence 21, Appl
47	139.5	3.3	5306	4	US-09-561-818A-21	Sequence 21, Appl
48	139.5	3.3	5330	3	US-09-562-702A-25	Sequence 25, Appl
49	138	3.3	6744	1	US-08-119-125A-2	Sequence 2, Appl
50	138	3.3	8982	3	US-08-976-255-5	Sequence 5, Appl
51	137.5	3.3	4833	4	US-09-513-783A-21	Sequence 21, Appl
52	137	3.3	6948	4	US-09-543-681A-1262	Sequence 1262, Appl
53	136.5	3.3	3827	2	US-08-447-031A-1	Sequence 1, Appl
54	136	3.3	2439	4	US-09-513-783A-3	Sequence 3, Appl
55	136	3.3	3171	4	US-09-513-783A-31	Sequence 31, Appl
56	136	3.3	3279	5	PCT-US93-03077-2	Sequence 2, Appl
57	135.5	3.3	4643	4	US-09-453-702B-22	Sequence 22, Appl
58	135.5	3.3	4899	2	US-08-210-535-5	Sequence 5, Appl
59	135	3.2	2582	2	US-08-816-105A-2	Sequence 2, Appl
60	135	3.2	2592	4	US-09-604-958-3	Sequence 3, Appl
61	135	3.2	5009	1	US-08-487-890A-3	Sequence 3, Appl
62	135	3.2	5009	2	US-08-478-435-3	Sequence 3, Appl
63	135	3.2	5009	2	US-08-337-483-3	Sequence 3, Appl
64	135	3.2	5009	2	US-08-337-483-3	Sequence 3, Appl
65	135	3.2	5009	3	US-08-474-671-3	Sequence 3, Appl
66	135	3.2	5009	3	US-08-483-577A-3	Sequence 3, Appl
67	135	3.2	5009	3	US-08-897-438-3	Sequence 3, Appl
68	135	3.2	5009	4	US-08-637-654-3	Sequence 3, Appl
69	135	3.2	5009	4	US-08-649-518-2	Sequence 3, Appl
70	135	3.2	5033	1	US-08-487-890A-2	Sequence 2, Appl
71	135	3.2	5033	2	US-08-478-435-2	Sequence 2, Appl
72	135	3.2	5033	2	US-08-337-483-2	Sequence 2, Appl
73	135	3.2	5033	3	US-08-478-373-2	Sequence 2, Appl
74	135	3.2	5033	3	US-08-474-671-2	Sequence 2, Appl
75	135	3.2	5033	3	US-08-483-577A-2	Sequence 2, Appl
76	135	3.2	5033	3	US-08-897-438-2	Sequence 2, Appl
77	135	3.2	5033	4	US-08-637-654-2	Sequence 2, Appl
78	135	3.2	5033	4	US-08-649-518-2	Sequence 2, Appl
79	134.5	3.2	4899	6	5268270-1	Patent No. 5268270
80	134.5	3.2	5194	1	US-09-599-652-1	Sequence 1, Appl
81	134.5	3.2	5194	2	US-08-642-846-1	Sequence 1, Appl
82	134.5	3.2	5194	4	US-09-264-604-1	Sequence 1, Appl
83	134.5	3.2	5194	4	US-10-002-389-1	Sequence 1, Appl
c 84	134.5	3.2	590073	4	US-08-545-5280-1	Sequence 1, Appl
85	134	3.2	6924	1	US-08-015-973-2	Sequence 2, Appl

86 134 3.2 6924 2 US-08-448-164-2 Sequence 2, Appli  
 87 134 3.2 9626 4 US-09-150-867-2 Sequence 2, Appli  
 88 133.5 3.2 4612 2 US-08-447-031A-8 Sequence 8, Appli  
 89 132 3.2 2817 4 US-09-620-312D-1085 Sequence 1085, Ap  
 90 132 3.2 4797 4 US-09-976-594-988 Sequence 988, App

## ALIGNMENTS

## RESULT 1

US-08-961-083-55  
 ; Sequence 55, Application US/08961083  
 ; Patent No. 6159469  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,083  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 55:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2389 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-961-083-55

Alignment Scores:  
 Pred. No.: 0 Length: 2389  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: 3 Gaps: 0

US-09-765-271-56 (1-796) x US-08-961-083-55 (1-2389)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20  
 DB 2 TCTTACGAGTTGGAGCTGATCAAGCTAGAACGGTTAAGGAAATAATCTGCTTTCCTAT 61  
 QY 21 IleAspGlyGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 DB 62 ATAGATGGAACCAAGCCGACGCAAAACCGAGAATTTGACTCTCTGATGAGGTTAGCAAG 121  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 DB 122 CGTGAAGGAATCAATCTGAGCAATTCGTCATCAAGATTAACAGACCAAGGCTATGTCAC 181

QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 DB 182 TCACATGCGGACCACTATCATTTACAAATGGTAAGGTTCTTATGACGCTATCATCAGT 241  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
 DB 242 GAAGAATTACTCATGAAAGATCCAAACTATAAGCTAAAAGATGAGGATATTGTTAATGAG 301  
 QY 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 DB 302 GTCAAGGCTGGATATGTTTATCAAGGTAGATGAAATAATCTATGTTTATCTTAAGGATGCT 361  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluLysAsnArgGlnLysGlnHisSer 140  
 DB 362 GCCCACGGGATAAGCTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 421  
 QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaAArgSerGln 160  
 DB 422 CAACATCGTGAAGGTGAAGTCCCAAGAAACAGATGGTGTGCTTGCCTTCGACGTTTCGCAA 481  
 QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 DB 482 GGACGCTATCTACAGATGATGTTTATATCTTTAANGCTTCTGATATCATAGAGATACT 541  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 DB 542 GGTGATGCTTATATCGTTCCTCATGGAGATCATTTACCATTCATCTTAAGAATCAGTTA 601  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 DB 602 TCAGCTAGCAGTTCGGCTGCTGCAGAGCCTTCCTATCTGTGTCGAGGAAATCTGTCAAT 661  
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 DB 662 TCAAGAACCTATCGCCGACAAATAGCGATTAACACTTCAAGAACAACTGGGTACTCTTCT 721  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 DB 722 GTAAGCAATCCAGGAACCTACAATACTAACACAAACAAACAGCAACACATAACAGTCAA 781  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 DB 782 GCAAGTCAAGGTAATGACATTTGATAGTCTCTTGAACACAGCTCTACAACTGCTTTGAGT 841  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 DB 842 CAACGACATGATAGATCTGATGGCCCTTGTCTTTGATCCAGCAACAATCACAAGTCAACA 901  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 DB 902 GCTAGAGGTGTTGCAGTGCACACGAGATCATTTACCACTTCCTTACTCTCAAAATG 961  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
 DB 962 TCTGAAATTTGGAAGAACGAAATCGCTGATTTATTCCTCTGCTTATTCGTTTCAACACATTGG 1021  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 DB 1022 GTACCAGATTTCAGGCCAGAACCAACAGTCCCAACCGACTCCGGAACCTAGTCCAGGC 1081  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 DB 1082 CGGCAACCTGCACCAAAATCTTAAAAATAGACTCAAAATCTCTCTTGTGTAGTCAGCTGTA 1141  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGlyLysGlyIleSerArgTyrValPheAla 400  
 DB 1142 CGAAAGTTGGGAAGGATATGATTCGAAGAAAAAGGGCACTCTCTGTTATGCTTTGGC 1201  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 DB 1202 AAAGATTACCATCTGAAACTGTTAAAAATCTTTGAAGCAAGTATCAAAACAAGAGAGT 1261  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440

Db 1262 GTTTCACACACTTTAACTGCTAAAAAGAAAATGTTGCTCTCGTGACCAAGATTTTAT 1321

Qy 441 AsplysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460

Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGNAATAAAGGTCGT 1381

Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480

Db 1382 AATTCTGATTTCCAGCCTTAGACAATATTAGAACGCTTGAATGATGAATCCACTAAT 1441

Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500

Db 1442 AAAGAAAAATTTGGTAGATGATTTATTTGGCAATCTCTAGCACCAATACCCATCCAGAGCGA 1501

Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520

Db 1502 CTTGGCAAAACCAATTTCAAAATTTGAGTATATCTGAAGACGAAGTTGCTTATGCTCAATTA 1561

Qy 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540

Db 1562 GCTGATAGATATACAACTGATGTTTACATTTTGTGATGAACATGATATATCAGTGAT 1621

Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560

Db 1622 GAAGCAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGATTTGGAAGAAGATAGC 1681

Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

Db 1682 CTTTCTCATAGGAAAAAGTTGACGCTCAACCTATATCAAGAAAAAGGATGCTACCT 1741

Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600

Db 1742 CCATCTCCAGCGCAGATGTTAAAGCAATCCAACTGGAGATAGTGACGAGCTATTATC 1801

Qy 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620

Db 1802 AATCGTGTAAAGGGAAAAACGAATTCACCTGCTTCGACTTCCATATATGTTGAGCAT 1861

Qy 621 ThrValGluValLysAsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsnIle 640

Db 1862 ACAGTTGAGGTTAAAAACGGTAATTTGATTTATTCCTCATAGGATCATTTACCATATATT 1921

Qy 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660

Db 1922 AATTTGCTTGTGTTGATGATCACATACAACTCCAAATGGCTATACCTTGGAAAGAT 1981

Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680

Db 1982 TTGTTTGGCAGCATTAAGTACTAGCTAGAACACACCTCGACCAACGTCACATTTCTAATGAT 2041

Qy 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700

Db 2042 GGATGGGGCAATGCCAGTGAGCATGTTGTAGGCAAGAAAGACCAAGATCCAAAT 2101

Qy 701 LysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720

Db 2102 AAGNACTTCAAGCGGATGAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 2161

Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740

Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTGTCTTGGCAAA 2221

Qy 741 ValThrAspSerSerLysLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760

Db 2222 GTAACCGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 2281

Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780

Db 2282 TTGACTCTTCAATTTGGATTAACAATAGTATATGCGCAGAGCAGAAAAATTTACTTGGC 2341

Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796

Db 2342 TTGTTAAAGGAAGTAATCTTCTCATCTGTAAGTAAGGAAAAATAAAC 2389

RESULT 2

US-09-536-784-55

; Sequence 55, Application US/09536784

; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/536,784

; FILING DATE: 30-Oct-1997

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/961,083

; FILING DATE: OCT-30-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971

; REFERENCE/DOCKET NUMBER: PB340P3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 55:

; LENGTH: 2389 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-536-784-55

Alignment Scores:

Pred. No.: 0 Length: 2389

Score: 4163.00 Matches: 796

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 99.95% Indels: 0

DB: 4 Gaps: 0

US-09-765-271-56 (1-796) x US-09-536-784-55 (1-2389)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20

Db 2 TCTTACGAGTTGGAGCTGTATCAAGCTAGAACCGTTAAGGAAAAATAATCGTGTTCCTAT 61

Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40

Db 62 ATAGATGGAANAACAGCGACGCAAAAACGAGAAATTTGACTCTCTGATGAGTTAGCAAG 121

Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60

Db 122 CGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGCTATGTCAT 181

Qy 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80

Db 182 TCACATGGCGACCATCATCATATTATTCAATGGTAAAGGTTCCCTTATGACGCTATCAT 241

Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100



```

; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2451)
; OTHER INFORMATION: n = a, c, t or g
US-09-468-656A-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 4 Gaps: 0

US-09-765-271-56 (1-796) x US-09-468-656A-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 61 TCTTACAGTGGGACTGTATCAAGCTAGAACCGTTAAGGAAAATAATCGTGTTCCTAT 120

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 121 ATAGATGGAAACNAGCAGCGAANAACGGAATTTTGACTCTGATGAGGTAGCAAG 180

QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 181 CGTGAAGGAATCAATGCTGAGCAATCGTCATCAAGATAACAGACCAAGCTATGTCAT 240

QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 241 TCACATGGCGACCACTATCATATTACAAATGGTAAGGTTCCTTATGACGCTATCATCAGT 300

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsnGlu 100
Db 301 GAAGAAATTAATCATGAAGATCCAACTATAAGCTAAAGATGAGGATATTGTTAATGAG 360

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 361 GTCAAGGGTGGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGCT 420

QY 121 AlaHisAlaAspAsnValArgThrLysGluLysIleAsnArgGlnLysGlnGluHisSer 140
Db 421 GCCCAGCGGATACGTCGCTACAAAGAGGAATCAATCAAGAAACGATGCTGTGCTGGCAGCTTCGCAA 480

QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyValAlaValAlaLeuAlaArgSerGln 160
Db 481 CAACATCGTGAAGGTGAATCCNAGAAACGATGCTGTGCTGTGCTGGCAGCTTCGCAA 540

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 541 GGACGCTATACACAGATGATGGTTATATCTTTAATGCTTCGATATCATAGAGGATACT 600

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 601 GGTGATGCTTATATCGTTCTCTCATGGAGATCATACCAATACATTCCTCAAGAATGAGTTA 660

```

```

201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
661 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCTATCTGCTGGAGGAAATCTGTCAAA 720

221 SerArgThrTyrArgAlaGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
721 TCAGAAACCTTATCGCCCAAAATAGGATACACTTCAAGAACAACTGGGTACTTCT 780

241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
781 GTAAGCAATCCAGGAATCTACAAATACAAACAGCAACACAGCAACACACTAACAGTCAA 840

261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
841 GCAAGTCAAAAGTAAATGACATTTAGTAGTCTCTTGAACACAGCTCTACAACTGCCCTTGGAT 900

281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
901 CAACGACATGTAGAAATCTGATGGCTTGTCTTTGATCCAGCAAAATCAACAGTCCGAA 960

301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
961 GCTAGAGGTGTTCAGTGCACACAGGAGATCATACCACTTCATCCCTTACTCTCAAAATG 1020

321 SerGluLeuGluAlaArgIleAlaAqlleProLeuArgTyrArgSerAsnHisTyr 340
1021 TCTGAATTTGAAGAACCAATCGCTCGTATTATTCCTCTTCTGTTATCTGTTCAAAACATTTG 1080

341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
1081 GTACCAAGTTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1140

361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
1141 CCGCAACTGCACCAAAATCTTAAATATAGACTCAAAATCTTCTTTGGTTAGTCAAGTGGTA 1200

381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
1201 CGAAAGTTGGGAAGGATATGATTTCGAAGAAAAGGCGATCTCTCTGTTATGCTTTTGGC 1260

401 LysAspLeuProSerGluThrValLysAsnLeuSerLysLysLeuSerLysGlnGluSer 420
1261 AAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAGT 1320

421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
1321 GTTTCACACACTTAACTGCTAAAAAAGAAATGTTGCTCTCTGTCGACCAAGAAATTTAT 1380

441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
1381 GATAAGCATATTAATCTGTTAACTGAGGCTCATAAAGCCTTGTITGNAAATAAGGGTCGT 1440

461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
1441 AATTTCTGATTTCAAGCCTTAGACAAATATTATGAAGCCTTTGAATGATGATGACTAAT 1500

481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500
1501 AAAGAAAAATTTGTAGATGATTTATTGGCATTCCTAGCACCAATTACCCATCCAGAGCGA 1560

501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValAlaGlnLeu 520
1561 CTTGGCAACCAAAATTTCTCAAATGAGTATCTGAAGACGAGTTCGTATGCTCAATTA 1620

521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
1621 GCTGATAAGTATACACAGCTCAGATGTTACATTTTGTATGATAACATGATAATCAAGTAT 1680

541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560
1681 GAAGGAGATGATATGTAAACGCTCATATGGGCCATAGTCACTGGATTTGAAAGATAGC 1740

561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

```



1741	Db	CTTTCTCATAGGAAGTTCAGCGCTCAAGCCCTATACAAAGAAAAGGTATCCTACCT	1800
581	Qy	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaLalleTyr	600
1801	Db	CCATCTCCAGACCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCGACGAGCTATTAC	1860
601	Qy	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
1861	Db	AATCGTGTGAAGGGGAAAAACGAATTCCTCGACTCCCATATATGGTTGAGCAT	1920
621	Qy	ThrValGluValLysAsnGlyAsnLeuLelleProHisLysAspHisTyrHisAsnIle	640
1921	Db	ACAGTTGAGGTTAAAAACGGTAATTTGATTATTCTCTCATAGGATCATTACCAATAATATT	1980
641	Qy	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
1981	Db	AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATATCCTTGGAGAT	2040
661	Qy	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
2041	Db	TTGTTTTCGCACGATTAAGTACTACGTAGAACACCCCTGACCAAGCTCCACATTTCTAATGAT	2100
681	Qy	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
2101	Db	GGATGGGGCAATGCGCAGTGAGCATGTGTGTAGGCAAGAAAGACCACACTGAGATGCCAAAT	2160
701	Qy	LysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro	720
2161	Db	AAGAACTTCAAGCGGATGAAGACCCAGTAGAGGAACACACTGCTGAGCCAGAGTCCCT	2220
721	Qy	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
2221	Db	CAACTAGAGACTGAAAAAAGTAGAAGCCCACTCAAGAAAGCAGAAAGTTTGTCTGCAAA	2280
741	Qy	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
2281	Db	GTAACGGATTCCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACCGAAATAAT	2340
761	Qy	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
2341	Db	TTGACTCTCTCAAAATATGGATAACAATAGTATCATGGCAGACGACGAAAAATTACTTGG	2400
781	Qy	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
2401	Db	TTGTTAAAGCAAGTAATCTCTTCATCTGTAGTAAGCAAAAAATAAC	2448

## RESULT 4

RES001 4  
 US-08-961-527-94  
 ; Sequence 94, Application US/08961527  
 ; Patent No. 6420135  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:

QY 241 ValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 3774 GTAAGCAATCCAGGAATACAAAATACAAACACAGCAACACAGCAACAGTCAA 3833  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLeuGlnLeuLeuLeuLeuSer 280  
 Db 3834 GCAAGTCAAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 3893  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 3894 CAACGACATGTAGAAATCTGATGGCTTCTGATGATGATGATGATGATGATGAT 3953  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisThrHisPheIleProTyrSerGlnMet 320  
 Db 3954 GCTAGAGTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4013  
 QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
 Db 4014 TCTGAATGGAGAACGATCGCTCGTATATATCCCTTCTGATGATGATGATGAT 4073  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 4074 GTACCAGATTCAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 4133  
 QY 361 ProGlnProAlaProAsnLeuLeuIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 4134 CCGACACTGCACCAATCTTAAATATAGACTCAAAATCTCTTGGTGTAGTCACTGGTA 4193  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 4194 CGAAAGCTGGGAGGATATGATTCGAAGAAAGGGCATCTCTCGTATGCTTTTGGC 4253  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 4254 AAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACCAAGAGAGT 4313  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 4314 GTTTCACACACTTTAACTGCTTAAAGAAAGAAATGTTGCTCTCGTACCAAGAAATTTAT 4373  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*AsnLysGlyArg 460  
 Db 4374 GATAAGCATATATATCTGTTAACTGAGCTCATTAAGCTTGTGTTGAAATTAAGGGTCTG 4433  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 4434 AATTCTGATTTCCAAGCCTTAGACAAATTAATAGAACGCTTGAATGATGAATCGACTAAT 4493  
 QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 4494 AAAGAAAATGGTAGATGATTTATGTCATCTCTAGCACAATACCCATCCAGAGCGA 4553  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 4554 CTTGGCAACCAATCTCAATTTGAGTATATCTAGACGAGTTCGATTTGCTCAATTA 4613  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 4614 GCTGATAGTATACACGTCAGTGGTGTACATTTTGTATGAACATGATATAATCAGTAT 4673  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560  
 Db 4674 GAAGAGATGATATGTAACCCCTCATATGGCCATAGTCTAGTGGTGGAAAGATAGC 4733  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 4734 CTTTCTGATAGGAAAAGTTGCGACTCAACCTATATCAAGAAAAGGATATCTACCT 4793  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 4794 CCATCTCCAGACGAGATGTTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTATC 4853

QY 601 AsnArgValLysGlyGluLysAspGlyProLeuValArgLeuProTyrMetValGluHis 620  
 Db 4854 AATCGTGTGAAGGGGAAAACGAATTCCTGCTGACTTCATATATGTTGAGCAT 4913  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle 640  
 Db 4914 ACAGTTCAGGTTAAACACGGTAAATTTGATTTATCTCTATAGGATCATACCATATATT 4973  
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 4974 AAATTTCTGTTGTTGATGATCACACATACAAAGCTCAAAATGGCTATACCTTCGAAGAT 5033  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 5034 TTGTTTCGAGGATTAAGTACTAGTAGAACCCCTCAGCAACCTCCACATTTCTATATGAT 5093  
 QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 5094 GGATGGGGCAATGCCAGTGAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAAAT 5153  
 QY 701 LysAsnPhelysAlaAspGluGluProValGluThrProAlaGluProGluValPro 720  
 Db 5154 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAAACACCTGCTGAGCCAGAAAGTCCCT 5213  
 QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740  
 Db 5214 CAAGTAGAGTCTGAAAAGTAGAAGCCCACTCAAGAACGAGAGTTTGTCTTGGCAAA 5273  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 5274 GTAACGGATCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGTTTACGAAATAAT 5333  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780  
 Db 5334 TTCACTCTTCAATTTATGATTAACAATAGTATCATGGCAGAAAGCAGAAAATTTACTTGC 5393  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 5394 TTGTTAAAGAGAGTAAATCTTCTATCTGTAAGTAAGGAAAAATAAAC 5441

RESULT 5  
 US-09-468-656A-11  
 ; Sequence 11, Application US/09468656A  
 ; Patent No. 6582706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-444  
 ; CURRENT APPLICATION NUMBER: US/09/468,656A  
 ; CURRENT FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-468-656A-11

Alignment Scores:  
 Pred. No.: 2,07e-267 Length: 2531  
 Score: 2786.50 Matches: 544  
 Percent Similarity: 77.06% Conservative: 84  
 Best Local Similarity: 66.75% Mismatches: 140  
 Query Match: 66.90% Indels: 47  
 DB: Gaps: 7

US-09-765-271-56 (1-796) x US-09-468-656A-11 (1-2531)

1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19  
 61 TCCTATGAGCTTGGAGTGTACCAAGCTGTCAGGATAAGAGAGTCTTAATCGAGTTGCT 120  
 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 121 TATATAGATGGTCATCAGCTCGTCAAAAGGAGGAGAACTTGACACAGATGAAGTCAGT 180  
 40 LysArgGluGlyLeuLeuAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 181 AAGAGGAGGGGATCAACCCCAACAAATGTTATCAAGATTACGGATCAAGGTTATGGT 240  
 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 241 ACCTCTCATGGAGACCAATATCATATATATATGGAAGGTTCCCTATGATGCCATCATC 300  
 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLysLeuLysAspGluAspIleValAsn 99  
 301 AGTGAAGAGCTCTCATGAAGATCCGAATATATCATGTTGAAGGATTTCAGACATTTGTCAT 360  
 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119  
 361 GAATCAAGGGTGGTATGTCATTAAGGTAACCGTAATACTATGTTTACCTTAAGGAT 420  
 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 421 GCRGCTCATGGGATAATATTCGACAAAGAGAGATTAAACGTCAGAGCAGGAAACGC 480  
 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 481 AGTCATATCATCAACTCA-----AGAGCAGATTAATGCTGTTGTCAGCCAGAGCC 531  
 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 532 CAAGACGTTATACACCGATGATGGTATATCTCAATGTCATCTGATATCAATGAGGAC 591  
 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 592 ACGGGTGATGCTTATGCTTCTCAGCGCAGCATTAACATTAACATTCCTAAGAATGAG 651  
 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 652 TTATCAGCTAGGAGTGTAGCTCTCAGAGAGCTTATGGAATGGG----- 696  
 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 697 -----AAGCAGGATCTCGCTCTCTCTCAAGTCTAGTTATATGCA 738  
 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGACTGTCTCACTCAACTTATCAT 798  
 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 799 CAA---AATCAAGGGGAAACATTTCAAGCCCTTTTACGTGAATTTGATGCTAAACCCCTTA 855  
 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 856 TCAGAACCCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGTCGA 915  
 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 916 ACCGCGCAGAGGTGATGCTGCTCATGTGTAACCATTAACCTTTATCCCTTATGAACAA 975  
 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 976 ATGCTGAATGGAAAGCAATGCTGCTATATATTCCTCTGTTATGTTTCAACACCAT 1035  
 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 1036 TGGGTACCAAGATTCAAGACCAAGAAACCAAGTCCACACACCGACTCCAGAACCTAGTCCA 1095  
 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379

1096 AGTCCGCAACCAAGCTCCCAAGCAATCAATTTGAT-----GGGAAATTGCTCAAGAAGCT 1149  
 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyLysSerArgTyrValPhe 399  
 1150 GTTCGAAAGTAGGCGATGGTTATGCTTTTGGAGGAATGGAGTTTCTCGTTATATCCCA 1209  
 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLeuSerLysLysSerLysGlnGlu 419  
 1210 GCCAAGGATCTTTACGAGAAACACAGCAGCGCATTTAGTACAAACTGCCAAGCAGGAA 1269  
 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439  
 1270 AGTTTATCTCATAGCTAGTAAGTAAAGAACTGACCTCCCATCTAGTGTATCGAGAAATTT 1329  
 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 1330 TACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTACTTGATATAAAGGT 1389  
 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuGluAlaGluAsnAspGluSerThr 479  
 1390 CGACAAGTTGATTTTGGCTTTGGATAACCTTTGGAACGACTCAAGGATGCTCAAGT 1449  
 480 AsnLysGluLysLeuValAspLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 1450 GATAAAGTCAAGTTAGTGAAGATATCTTGCTTCTTAGCTCCGATCTGTCATCCAGAA 1509  
 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519  
 1510 CGTTAGGAAACCAAAATGCGCAATTTACCTACCTGATGATGAGATCAAGTAGCCAAAG 1569  
 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539  
 1570 TTGGCAGCAAGTACACACGACGAGCGGTATATCTTTGATCTCTGCTGATATAACAGT 1629  
 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559  
 1630 GATGAGGGGATGCTTATGTAACCTACATATGACCCATGAGCCTGATGATTAAGAAAGAT 1689  
 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579  
 1690 AGTTTGTCTGAGAGTGAGAGCGCGCAGCCAGGCTTATGCTTAAGAAAGAGTTTGACC 1749  
 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599  
 1750 CTTCTCTTCGACAGACCATCAGGATTCAGGAATATCTGAGGCAAAAGGACGACAGCTATC 1809  
 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619  
 1810 TACAACCGMTGAAAGCAGCTAAGAAAGGTGCCACTGTATGCTGCTTACATCTTCAA 1869  
 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639  
 1870 TATACCTAGAGAGTCAAAACCGGTAGTTAATCATACCTCATATGACCATTAACATAAC 1929  
 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659  
 1930 ATCAATTTGAGTGGTTTTCAGCAAGGCTTTATGAGGCACCTTAAGGGGTATATCTTGAG 1989  
 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679  
 1990 GATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAACGATCCGATTCAGT 2049  
 680 AspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699  
 2050 AATGTTTTGTTAAGCTAGCGACCATGTTTCAAGAAACAAATAATGCTCAAGCTGATACC 2109  
 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713  
 2110 AATCAACGGAACCAAGCAGGAGGAAACCTCAGACAGAAAAACCTGAGGAGAAACCC 2169  
 713 ----- 713

Db 2170 CCTCGAGAGAGAAACCGAAACGAGAACACAGAGTCTCCAAACCAACAGAGAGAACCA 2229  
 Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730  
 Db 2230 GAAGAAATCACAGAGGAATCAGAGAAACCTCAGGTCGAGACTGAAAGAGTTGAAGAAAAA 2289  
 Qy 731 LeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750  
 Db 2290 CTGAGAGAGGCTGAGATTTACTTGGAAATAATCCAGATCAATATCAAGTCCAAATGCC 2349  
 Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnSer 770  
 Db 2350 AAAGAGACTCTCAGAGGATTAAAAATAATTTACTATTGGCACCAGGACCAACAATACT 2409  
 Qy 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785  
 Db 2410 ATTATGGCAGAGCTGAAATACTATTGGCTTTATTAAAGGAGACT 2454

## RESULT 6

US-09-468-656A-5  
 ; Sequence 5, Application US/09468656A  
 ; Patent No. 6582706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-444  
 ; CURRENT APPLICATION NUMBER: US/09/468,656A  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-468-656A-5

## Alignment Scores:

Pred. No.: 5,75e-266 Length: 2531  
 Score: 2772.00 Matches: 541  
 Percent Similarity: 75.78% Conservative: 91  
 Best Local Similarity: 64.87% Mismatches: 142  
 Query Match: 66.55% Indels: 60  
 DB: 4 Gaps: 8

US-09-765-271-56 (1-796) x US-09-468-656A-5 (1-2531)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 Db 61 TCTATGAACCTTGGTCTGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 120  
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrAlaLeuThrProAspGluValSer 39  
 Db 121 TATATAGATCGTATCAGGCTGGTCAAAAGCAGAAACCTTGACCCAGATGAAGTCAGT 180  
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db 181 AAGAGGAGGGGATCAACGCCGGAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTG 240  
 Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIle 79  
 Db 241 ACCTCTCATGGAGACCAATTATCATTTATAATGGCAAGGTCCTTATGATGCCATCATC 300  
 Qy 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db 301 AGTGAAGAGCTCTCATGAAGATGCCAATATATCAGTTGAAGGATTCAGACATTTGTCAAT 360  
 Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119

Db 361 GAATCAAGGGTGGTTATGTTATCAAGGTAGATCGAAATACTATGTTTACCTTAAGCAT 420  
 Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 Db 421 CGAGCTCATCGGATAATATTCGACAAAGAGAGATTAAACGTCAGAACGAGAACAC 480  
 Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db 481 AGTCATTAATCAAGGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGACGAGAGCC 534  
 Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db 535 CAAGGACGCTATACACGAGATGATGTTATATCTTCAATGCATCTGATATCATTTAGGAC 594  
 Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db 595 ACGGGTGATGCTTATATCGTTCTCTCAGGGGACCATTAACCATTAACATTCCTAAGAATGAG 654  
 Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 655 TTATCAGCTAGCAGATTAGCTGCTGCAGAAAGCCTATTGGAAATGGG----- 699  
 Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 Db 700 -----AAGCAGGGATCTCGTCTCTTCAAGTCTCTAGTTATAATGCA 741  
 Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 742 AATCCAGCTCAACCAAGATTGTGACAGAACCAACAATCTGACTGTCACTCCAACTTATCAT 801  
 Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db 802 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTAATTGTATGCTAAACCCCTTA 858  
 Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db 859 TCAGAACGCCATGTGGAATCTGATGGCTTTATTTTCGACCAGCGCAATACACAGTCGA 918  
 Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db 919 ACCGCCAGAGGTAGTGTCTCTCATGGTAAACCATTAACCATTTATCTCTTATGAACAA 978  
 Qy 320 MetSerGluLeuGluAlaArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 Db 979 ATGTCTGAATTGGAAACCAAGATTGCTCGTATTATTCCCTTCGTATCTGTTCAAAACCAT 1038  
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1039 TGGGTACCAGATTCAGACCCAGAACCAACCAAGTCCCAANTCGACTCCGAAACCTAGTCCA 1098  
 Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 Db 1099 AGTCGCGCACTCGACCAAAATCTCAACCAAGCTCCCAAGCAATCCAATTGAT-----GAG 1152  
 Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
 Db 1153 AAATTGCTCAAGAGCTGTTTCGAAAGTAGGCGATGGTTATGCTTTGAGGAGAAATGGA 1212  
 Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 Db 1213 GTTCTCTGTTATATCCAGCCCAAGGATCTTTCAGCAGAAACACAGCAGCAGCATTTGATAGC 1272  
 Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 Db 1273 AAATGCCCAAGCAGGAAAGTTTATCTAAGCTAGGAGCTAAGAAACCTGACCTCCCA 1332  
 Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 Db 1333 TCTAGTGTGAGAAATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGAT 1392  
 Qy 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
 Db 1393 TTACTTGATAATAAAGTTCGACCAAGTTGATTTTGAGGCTTTGGATACCTGTTGGAAACGA 1452

RESULT 7		US-08-961-083-65	
		; Sequence 65, Application US/08961083	
		; Patent No. 6159469	
		; GENERAL INFORMATION:	
		; APPLICANT: Choi et. al.	
		; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines	
		; NUMBER OF SEQUENCES: 452	
		; CORRESPONDENCE ADDRESS:	
		; ADDRESSEE: Human Genome Sciences, Inc.	
		; STREET: 9410 Key West Avenue	
		; CITY: Rockville	
		; STATE: Maryland	
		; COUNTRY: USA	
		; ZIP: 20850	
		; COMPUTER READABLE FORM:	
		; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage	
		; COMPUTER: HP Vectra 486/33	
		; OPERATING SYSTEM: MSDOS version 6.2	
		; SOFTWARE: ASCII Text	
		; CURRENT APPLICATION DATA:	
		; APPLICATION NUMBER: US/08/961,083	
		; FILING DATE:	
		; CLASSIFICATION: 435	
		; PRIOR APPLICATION DATA:	
		; APPLICATION NUMBER:	
		; FILING DATE:	
		; ATTORNEY/AGENT INFORMATION:	
		; NAME: Brookes, A. Anders	
		; REGISTRATION NUMBER: 36,373	
		; REFERENCE/DOCKET NUMBER: PB340P2	
		; TELECOMMUNICATION INFORMATION:	
		; TELEPHONE: (301) 309-8504	
		; TELEFAX: (301) 309-8512	
		; INFORMATION FOR SEQ ID NO: 65:	
		; SEQUENCE CHARACTERISTICS:	
		; LENGTH: 2290 base pairs	
		; TYPE: nucleic acid	
		; STRANDEDNESS: double	
		; TOPOLOGY: linear	
		US-08-961-083-65	
		Alignment Scores:	
		Pred. No.: 7,82e-254 Length: 2290	
		Score: 2649.50 Matches: 519	
		Percent Similarity: 76.51% Conservative: 77	
		Best Local Similarity: 66.62% Mismatches: 130	
		Query Match: 63.61% Indels: 53	
		DB: 3 Gaps: 8	
		US-09-765-271-56 (1-796) x US-08-961-083-65 (1-2290)	
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal	--LysGluAsnArgValSer 19
DB	5	TCCTATGAACTTGGTCCTCACCAGCTGGTCAGGTTAGAAAGAGTCTAATCGAGTTCT	64
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer	39
DB	65	TATATAGATGGTGATCAGCTGGTCAAAAGGCAGAAACTTCACACCCAGATCAAGTCACT	124
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59
DB	125	AAAGAGGAGGGGATCAACGCCCAAAATNGTNAATCAAGATTACCGATCAAGTTATGTG	184
QY	60	ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle	79
DB	185	ACCTCTCATGGAGACCATATATCTATATATGCAAGGTTCCCTATGATGCCATCATC	244
QY	80	SerGluGluLeuLeuMetIysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99
DB	245	ACTGAAGAGCTCCTCATGAAGATCCGAATTTATCATGTTGAAGGATTCAGACATTGTCAAT	304
QY	100	GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119

```

Db 305 GAAATCAAGGTTGTTATGTCATTAAGTAAACGGTAAATCTATGNTACCTTAAGAT 364
Qy 120 AlaAlaHisAspAsnValArgThrLysGluGluLeuAenArgGlnLysGlnGluHis 139
Db 365 GCAGCTCATGCGGATAATATCGGACAAAAGAGAGATTAACGTCAGAACGAGCAACG 424
Qy 140 SerGlnHisArgGluGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 425 AGTCATTAATCATNACTCA-----AGACGAGATAATGCTGTGTGCGACGAGAGCC 475
Qy 160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 476 CAAGGAGCTTATACACGAGTATGGGTATATCTTCAATGATCATCTGATATCATTTAGGAC 535
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 536 ACGGGTGATGCTTATATCGTTCCTTCAGCGGACCATTAACCATTAACATTCCTTAAGAATGAG 595
Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 596 TTATCAGCTACGAGTATGCTGTCGAGAACCTTATTGGAATGGG----- 640
Qy 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db 641 -----AAGCAGGATCTGCTCTTCTCAAGTCTAGTTATAATGCA 682
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 683 AATCCAGCTCAACAGATGTCAGAGAACCCACAAATCTGAGTCACTCCCAACTTATCAT 742
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 743 CAA---AATCAAGGGAAACATTTCAAGCCCTTTTACGTGAATGTATGCTAAACCCCTTA 799
Qy 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 800 TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCAGCCAGCGCAATCACAGTCGA 859
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 860 ACCGCCAGAGGTGATGTCCTCATGCTAAACCATTAACCATTTATCTCCCTTATGAACAA 919
Qy 320 MetSerGluLeuGluArgGlyIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
Db 920 ATGTCTGAATGGAAACGAATTCGCTGATTAATCCCTTCCTTCTGTTATPGTTCAAACCAT 979
Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db 980 TGGGTACCATTCACAGACCAAGAACCAAGTCCACATCGACTCCGGAACCTAGTCCA 1039
Qy 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373
Db 1040 AGTCCGCAACCTTGACCAAAATCTCAACACGACTCCAGCAATCAATCAATTTGAT-----GAG 1093
Qy 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393
Db 1094 AAATGTCTAAGAGCTGTTCGAAAGTAGCGGATGTTATGCTTTTGAGGAGAAATGGA 1153
Qy 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413
Db 1154 GTTTCCTGTATATCCCGACCAAGATCTTCACAGAGAACAGCAGCAGGATTTGATGAGC 1213
Qy 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433
Db 1214 AAATGCGCAAGCAGGAGAAAGTTTATCTCATAGTACGAGCTAAGAAAACCTGACCTCCA 1273
Qy 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453
Db 1274 TCTAGTATCGAGAAATTTTCAATTAAGCTTTATGACTTACTAGCAAGAAATTCACCAAGAT 1333
Qy 454 LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473

```

```

Db 1334 TTACTTGATTAATAAAGTCGACAGTGTGATTTTGGAGCTTTGGATAACCTGTTGGAACGA 1393
Qy 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493
Db 1394 CTCAGGATGTGTCNCAAGTGTAAAGTCAAGTGTAGTGGANGATAATCTTGCTCTTCTAGCT 1453
Qy 494 ProfileThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513
Db 1454 CCGATTTCGTATCCAGAACGTTTAGGAAAACCAATGCGAAAATACCTACACTGATGAT 1513
Qy 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
Db 1514 GAGATTCAAGTAGCAAGTTTGGCAGGCAAGTACACACAGAGACGGTATATCTTTGAT 1573
Qy 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
Db 1574 CCTCGTATATAACAGTATGAGGGGATGCTTATGTAATCCACATATATGACCATAGC 1633
Qy 554 HistripIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573
Db 1634 CACTGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGCGGCGAGCCCGAGGCTTATGCT 1693
Qy 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
Db 1694 AAAGAGAAAGTTTGACCCCTCTTCGACAGACCATCAGGATTTCAGGAAATACTGAGGCA 1753
Qy 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613
Db 1754 AAAGGAGCAAGAGCTATCTACACCGCTGAAAGCAGCTAAGAGGTGCCACTTGATCGT 1813
Qy 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633
Db 1814 ATGCTTACATCTTCAATATCTGTAGAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1873
Qy 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
Db 1874 TATGACCATTTACCATATACATCAATTTTGTAGTGTTCAGGAGGCCCTTTATGAGGCACCT 1933
Qy 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
Db 1934 AAGGGGTATCTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 1993
Qy 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
Db 1994 GAACGTCGCGCATTCAGATAATGTTTGTGTAACGCTAGCAGCATCTGTCAAGAAAACAA 2053
Qy 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
Db 2054 AATGTCACAGCTGATACCAATCAACCGAAAACCAAGCGAGGAGAAACCTCAGACAGAA 2113
Qy 709 ---ProValGluGluThr----- 713
Db 2114 AAACCTGAGGAAGAAACCCCTCGAGAGAGAAACCGCAAGCGAGAACCCAGAGTCTCCA 2173
Qy 714 -----ProAlaGluProGluValProGlnValGluThr 724
Db 2174 AAACCAACAGAGAACACAGAGAAATCCAGAGGAATCAGAGAACCTCAGTCCAGACT 2233
Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743
Db 2234 GAAAAGGTTGAAGAAAACCTGAGAGAGGCTGAAGATTACTTGGAAAATATCCAGAT 2290

```

## RESULT 8

US-09-536-784-65

; Sequence 65, Application US/09536784

; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2290 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 65:  
US-09-536-784-65

Alignment Scores:  
Pred. No.: 7,82e-254 Length: 2290  
Score: 2649.50 Matches: 519  
Percent Similarity: 76.51% Conservative: 77  
Best Local Similarity: 66.62% Mismatches: 130  
Query Match: 63.61% Indels: 53  
DB: 4 Gaps: 8

US-09-765-271-56 (1-796) x US-09-536-784-65 (1-2290)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnArgValSer 19  
DB 5 TCCTATGAACCTGGTCGTCAACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTCT 64  
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
DB 65 TATATAGATGGTGATCAGCTGGTCAAAAGGCGAGAAACTTGACACCAAGTCAAGTCAGT 124  
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
DB 125 AAGAGGAGGGGATCAACCCCAACAAATNGTNATCAAGATTACGATCAAGGTTATGTG 184  
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIlelle 79  
DB 185 ACCTCTCATGGAGACCATATCATCTACTATTAATGCAAGGTTCTTATGATCCCATCATC 244  
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
DB 245 AGTGAAGAGCTCCTCATGAAGATCCGAATTATCATGTTGAAGATTTCAGACATTGTCAT 304  
QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
DB 305 GAAATCAAGGTTGGTTATGTCATTAAGGTAACCGTTAATATACTATCTATACCTTAAGGAT 364  
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
DB 365 GCAGCTCATCGGATAATATTCGACAAAGAGAGATTAAACGTCAGAGAGGAGAAACGC 424  
QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
DB 425 AGTCATAATCAAACTCA-----AGAGCAGATATATGCTGTGCTCGACCGAGGCC 475

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
DB 476 CAAGACGTTTATACAAACGATGGTATATCTTCAATGATCATCATGATCATCATGAGGAC 535  
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
DB 536 ACGGGTGATGTTATATCGTTCCTCACGGGACCATTACCATTAATCATCTTAAGAATGAG 595  
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
DB 596 TTATCAGCTAGCGAGTTAGCTGCTGCAGAACGCTTATGGAATGGG----- 640  
QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
DB 641 -----AAGCAGGATCTCGTCTCTTCAAGTCTAGTTATATATGCA 682  
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
DB 683 AATCCAGCTCAACCAAGATTGTGAGAACCAACATCTGACTGTCTCACTCAACTTATCAT 742  
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
DB 743 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTTAAACCCCTTA 799  
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
DB 800 TCAGAACCCCATGTGAATCTGATGCGCTTATTTGACCCAGCCCAATCAAGTCA 859  
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
DB 860 ACCGCCAGAGGTGTAGCTGCTCCTCATGTAACCATTACCATTTATCCCTTATGAACAA 919  
QY 320 MetSerGluLeuGluAlaArgIleAlaArgIleIleProLeuArgTyrArgSerHis 339  
DB 920 ATGCTGAATTGGAATAACGAATTCTCGTATTATTCCTCTTCGTTATGCTTCAACCAT 979  
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
DB 980 TGGGTACAGATTCAAGACCAACCAACAGTCCACAACTCGACTCCGGAACCTAGTCCA 1039  
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
DB 1040 AGTCGGCAACCTGCACCAATCTCAACAGCTCCCAAGCAATCCAATTCAT-----GAG 1093  
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
DB 1094 AAATTTGGTCAAGAAAGCTGTTGCAAAAGTAGCGATGTTATGCTTTTGAAGGAGATGGA 1153  
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
DB 1154 GTTCTCTGTTATATCCAGCCAGGATCTTTCAGCAGAAACAGCAGCAGGCTTATGATAGC 1213  
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysGlyLysValAla 433  
DB 1214 AAATGGGCAAGCAGGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACCTACACCTCCA 1273  
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAla 453  
DB 1274 TCTAGTGTATCAGAAATTTTACAAATAGGCTTATGATTTACTAGCAGAAATTCACCAAGAT 1333  
QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg 473  
DB 1334 TTACTTGATATAAAGGTCGACAAAGTGTATTTTGAGGCTTTTGATAAACCCTGTGGAACGA 1393  
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
DB 1394 CTCAGAGGATGTCNCAGGTGATAAAGTCAAGTTAGTGGAGATATCTTTCCTCTTAGCT 1453  
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
DB 1454 CCGATTCTCATCCAGAACGTTTAGAAAAACCAATGCGCAATTACCTACACTGATGAT 1513



```

QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533
DB 1514 GAGATTCAAGTAGCCAGGTGGCAGGACGATACAAACAGAACGGTTATATCTTTGAT 1573
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553
DB 1574 CCTCGTGATATAACCAAGTAGGAGGGGATGCTATGTAACTCCACATATGACCCATAGC 1633
QY 554 HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThr 573
DB 1634 CACTGGATTAAAAAGATAGTTCTCTGAAGCTGAGAGCGGCGAGCCAGGCTTATGCT 1693
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593
DB 1694 AAAGAGAAAGTTTGACCCCTTCCTTCGACAGACCATCAGGATTCCAGAAATACCTGAGGCA 1753
QY 594 AspSerAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613
DB 1754 AAAGAGCAGAGAGCTATCTCAACCGCGTGAAGCAGCTAAGAAGGTGCCACTTGATCGT 1813
QY 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleProHis 633
DB 1814 ATGCTTTACAACTTCTCAATATCTGTAGAGTCAAAACCGTAGTCTTAATCATACCTCAT 1873
QY 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653
DB 1874 TATGACCATTTACATAACATCAATTTGAGTGGTTGACGAGGCCCTTTATGAGGCACCT 1933
QY 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673
DB 1934 AAGGGGTATCTCTTGAGGATCTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC 1993
QY 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLys 693
DB 1994 GAACGTCGCGATTACAGATAATGTTTGGTAAACGCTAGCGACCATGTTCAAAGAAACAAA 2053
QY 694 AspHisSerGluAspProAsnLysAsnPheLysAlaAspGluGlu----- 708
DB 2054 AATGGTCAAGCTGATACCAATCAACCGAAACCAAGCGAGGAGAAACCTCAGACAGAA 2113
QY 709 ---ProValGluGluThr----- 713
DB 2114 AAACCTGAGGAAGAAACCCCTCGAAGAGAAACCGCAAGAGAGAAACCGAGAGTCTCCA 2173
QY 714 -----ProAlaGluProGluValProGlnValGluThr 724
DB 2174 AAACCAACAGAGAAACCGAGAAATCCACGAGGAATCAGAGAAACCTCAGGTCGAGACT 2233
QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAsp 743
DB 2234 GAAAGGTTGAAGAAACCTGAGAGAGGCTGAAGATTACTTGGAAATAATCCAGGAT 2290

```

## RESULT 9

```

US-08-961-527-243
; Sequence 243, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-243

```

## Alignment Scores:

```

Pred. No.: 1.37e-166 Length: 2359
Score: 1774.00 Matches: 347
Percent Similarity: 81.22% Conservative: 51
Best Local Similarity: 70.82% Mismatches: 74
Query Match: 42.59% Indels: 18
DB: 4 Gaps: 5

```

US-09-765-271-56 (1-796) x US-08-961-527-243 (1-2359)

```

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
DB 940 TCCTATGAGCTTGAGCGTTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 999
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
DB 1000 TATATAGATGGTGATCAGGCTGGTCAAAAGCGAGAAACTTGACACGATGAAGTCAGT 1059
QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
DB 1060 AAGAGGGAGGGGATCAACGCCGAAACAAATTGTTTCAAGATTACGGATCAAGGTTATGTG 1119
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
DB 1120 ACCTCTCATGGAGACCATTTATCTATATATGCGCAAGGTTCTTATGATGCCATCATC 1179
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
DB 1180 AGTGAAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGGATTCAGACATTGTCAAT 1239
QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
DB 1240 GAAATCAAGGGTGGTTATGTTCATTAAAGTAAACGGTAATACTATGTTTACCTTAAGGAT 1299
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
DB 1300 GCAGCTCATCGGATTAATATTCGACAAAAGAGAGATTAACAGTCAAGACGAGCAACGC 1359
QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
DB 1360 AGTCATAATCATAACTCA-----AGACGAGATAATGCTGCTGCGAGCCAGAGCC 1410
QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
DB 1411 CAAAGGAGGTTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTGAGGAC 1470
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
DB 1471 ACGGGTGATGCTTATATCGTTCTCAGCGGACCATTAACCATTACATTCTTAAGATGAG 1530
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219

```

Db 1531 TTATCAGCTAGCAGGTTAGCTGCTGACAGAGCTATTGGAATGGG----- 1575  
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
1576 -----AAGCAGGAGTCTCGTCTCTTCAAGTCTAGTATAATGCA 1617  
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
1618 AATCAGCTCAACAGAGATTGTCAGAGAACACAACTGCTGCTCCTCACTCCACCTATCAT 1677  
Qy 260 GlnAlaSerGlnSerAsnAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
1678 CAA---AATCAAGGGGAAACATTTCAAGCTTTTACGCTGTAATGCTGCTAAACCTTA 1734  
Qy 280 SerGlnArgHisValGlnSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
1735 TCAGAACCCATGTCGATGCTGCTGCTTATTTTCGACCCAGCCAAATCACAAGTGA 1794  
Qy 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
1795 ACCGCCAGAGTGTAGCTGCTCCTCATGTTACCACTTACCCTTTATCCCTTATGAACA 1854  
Qy 320 MetSerGluLeuGluGluArgGlnAlaGlnIleProLeuArgTyrArgSerAsnHis 339  
1855 ATGCTGAATTTGGAAGAAACGAATTCGCTGCTATTTATCCCTTCGTTATCGTTCAACAC 1914  
Qy 340 TrpValProAspSerArgProGlnProGlnProSerProGlnProThrProGlnProSerPro 359  
1915 TGGGTACCAGATTTCAAGACCAAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1974  
Qy 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379  
1975 ACTCGCAACAGCTCCAGCAATCCATTTGAT-----GAGCAANTTGTCAAGAGCT 2028  
Qy 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399  
2029 GTTCGAAAGTAGTGGCGATGTTATGCTTTGAGGAGAAATGAGATTTCTGTTATATCCA 2088  
Qy 400 AlaLysAspLeuProSerGlnThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
2089 GCCAAGGATCTTCAGCAAGAAACAGCAGCGCATTTGATGCAAACTGGCCCAAGCAGGAA 2148  
Qy 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439  
2149 AGTTATCTCATAGCTAGGAACCTAAGAACTGACCTCCCATCTAGTATGATGAGATTT 2208  
Qy 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
2209 TACAATAAGGCTTATGACTTACTAGCAGAATTCACCAAGATTTACTTGAATAAAGGT 2268  
Qy 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
2269 CCACAAGTTGATTTTGGAGCTTTGGAAACCTTTGGAAACGCTCAAGATGCTCAAGT 2328  
Qy 480 AsnLysGluLysLeuValAspAspLeuLeu 489  
2329 GATAAAGTCAAGTTAGTGAAGATATCTTT 2358

## RESULT 10

US-09-765-271-56

Sequence 192, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-192

Alignment Scores:  
Pred. No.: 6,766-111 Length: 6867  
Score: 1222.50 Matches: 295  
Percent Similarity: 54.22% Conservative: 97  
Best Local Similarity: 40.80% Mismatches: 181  
Query Match: 29.35% Indels: 151  
DB: 4 Gaps: 18

US-09-765-271-56 (1-796) x US-08-961-527-192 (1-6867)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
Db 6722 GCCTATGCATTAACCCAGCATCGTTTCGAGGAGAAATAAGGACATAATCGTCTCTTAT 6663  
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
Db 6662 GTGGATGCGACGACCACTCAAGTCAGAAAAGTGAACCTTGACACCCAGACGTTAGCCAG 6603  
Qy 41 ArgGluGlyIleAsnAlaGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
Db 6602 AAAGAAGGAATTCAGGCTGAGCAAAATTTGATCAAAATTCAGATCAGGCTATGTACG 6543  
Qy 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleSer 80  
Db 6542 TCACACGGTGACCATCTATCTATTAATGGGAAAGTTCTTATGATGCCCTCTTATG 6483  
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
Db 6482 GAAGAACCTCTTGATGAAGATCCAACTATCACTTAAAGACGCTGATATGTCATGAA 6423  
Qy 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAspAla 120  
Db 6422 GTCAAGGGTGGTTATATCATCAAGTCGATGGAATAATTTATGCTACTCTGAAGATGCA 6363  
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140  
Db 6362 GCTCATGCTGATAATGTTTCGAACCTAAAGATGAATAATCAATCGTCAAAAACAAGACATGTC 6303  
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
Db 6302 AAAGATAATGAG-----AAGGTTAACTTAATGTTGCTGTCAGAGGCTCTG 6255  
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
Db 6254 GGACCATATACGCAAAATGATGTTTATGCTTTTAAATCCAGCTGATATTCGAAGATACG 6195

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 DB 6194 GGTAAATGCTTATATGCTTCTCATGAGCTCACTATCACTATCCCAAGCAATTA 6135  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 DB 6134 TCTGCTAGTGAATAGCAGCAGCTAAAGCACAATCTGGCTGGAAAAAATATGCAACCGAGT 6075  
 QY 221 SerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 DB 6074 CAGTTAGCTATTCTTCAACAGCTAGTCAAT 6042  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 DB 6041 -----AACAGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCA 6003  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLysSer 280  
 DB 6002 GCAATAAATCTGAATAATCTCCAGAGCTCTTTGAAGGAACCTATGATTCACCTAGCGCC 5943  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 DB 5942 CAACGTTACAGTGAATCAGATGGCTGTCTTTGACCTGCTAGATATCAGTCGTACA 5883  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 DB 5882 CCAATGGAGTGGCAATTCGCATGGCGACCACTTACCACCTTTATTCCTTACAGCAAGCTT 5823  
 QY 321 SerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340  
 DB 5822 TCTGCCTTAGAAGAAAGATTGCGAGATGGCGCT 5787  
 QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 DB 5787 ----- 5787  
 QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 DB 5786 -----ATCAGTGAAGTGGTCTTACAGTTTCTACAAATGCA 5751  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 DB 5750 AAA----- 5748  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnGluSer 420  
 DB 5747 -----CCTAATGAAGTAGTGTCTAGCTAGCGAGCTTTTCAAGCAATCCTCTCTCT 5697  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 DB 5696 -----TTAAGCACAAGTAAGGAGCTCTCTTCCAGCATCTGATGTTATATTTT 5649  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 DB 5648 AATCCAAA-----GATATCGTTGAAGAACCGGTACAGCTTATATGTAAGACATGCTG 5595  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 DB 5594 ---GATCAATTCATCACTCCAAA-----TCAATCAAAATGGG 5556  
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 DB 5555 CAACCGACTCTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAAT 5499  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 DB 5498 CAGGAACCTTCAATGAGAAACATGAA----- 5472  
 QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 DB 5471 -----GNAAGATGATACGGAATTTGATGCTAATCGTATATTCGCTGAA 5430  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560

DB 5429 GATGAATCAGTTTTTGTCTAGTCCAGGACCAATCATTTATTTCTTCAAGAAAGAC 5370  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 DB 5369 TTGACAGAAGACAAATTA--GSTGCGCAAAACATTTAGAGAA----- 5326  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 DB 5325 -----GTTAAACCTAGTCAATATGATTTAGATTCTTTGTCTATCTCAT 5284  
 QY 601 AsnArg-----ValLysGlyGluLys 607  
 DB 5283 GAACAGGATTATCCAAAGTAATGCCAAGAAATGAAGATTAGATAAAAAAATCGAAGAA 5224  
 QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625  
 DB 5223 AAAATTGTGCAATTATGAAACAATATGCTGCTCAACCTGAAAGTATTGTCTGTAATAA 5164  
 QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTyr 644  
 DB 5163 GAAAAAATGCGATTATTTATCCGCTGAGATCACCATCATCGATCCG----- 5113  
 QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664  
 DB 5112 ATTGATGAACAT-----AAACCGTTGGAATTGGT-----CATTTCCAC 5074  
 QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681  
 DB 5073 AGTAACATGAACTGTTTAAACCGAAGAGGAGTTCCTAAAAAGAGGAATAAAGTT 5014  
 QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 DB 5013 TATACTGGAGAAGATTACGAATGTTTAAATTTGTTAAAAAATAGTACGTTTAAATAAT 4954  
 QY 701 LysAsnPhe 703  
 DB 4953 CAAACCTT 4945  
 RESULT 11  
 US-09-468-656A-7  
 ; Sequence 7, Application US/09468656A  
 ; Patent No. 6582706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-444  
 ; CURRENT APPLICATION NUMBER: US/09/468,656A  
 ; PRIOR FILING DATE: 1999-12-02  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 1455  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-468-656A-7  
 Alignment Scores:  
 Pred. No.: 4.28e-110 Length: 1455  
 Score: 1203.50 Matches: 259  
 Percent Similarity: 59.47% Conservative: 77  
 Best Local Similarity: 45.84% Mismatches: 120  
 Query Match: 28.90% Indels: 109  
 DB: 4 Gaps: 10  
 US-09-765-271-56 (1-796) x US-09-468-656A-7 (1-1455)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20

```
Db 64 GCCATGACATTAACACGATCGTTCGAGAAATAAGCAATAATCGTCTCTAT 123
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGGATGCGCAGCAGTCAAGTCAGAAAAGTGAACACTTGACACACAGACGAGTACG 183
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60
Db 184 AAAGAAGAAATTCAGCTGAGCAAAATGTAATCAAAATTTACAGATCAGGCTGATTAACG 243
Qy 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 244 TCACAGGTGACCATCATCTATCTATTAATGGAAAGTTCCTTATGATGCCCTCTTTAGT 303
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 304 GAAGAACTCTTGATGAAGATCCAACTATCAACTTAAAGACGCTGATATGTCATGA 363
Qy 101 ValLysGlyGlyTyrValLysValLysValAspGlyLysTyrValTyrLysLysAspAla 120
Db 364 GTCAAGGTGGTATATCATCAAGGTGATGGAATATATGTTCTCTCTGAAAGATGCA 423
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer 140
Db 424 GCTCATGCTGATATGTTGCAACTAAGATGAATCAATCGTCAAAACAAAGACATGTC 483
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 484 AAAGATAATGAG-----AAGGTTAACTCTAATGTTGCTGTAGCAAGGTCCTAG 531
Qy 161 GlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 532 GGACATATACCAACAAATGATGTTATGTTTATCCAGCTGATATATCAAGATACG 591
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAAATGCTTATATCGTCTCATGGAGGTCACATCACTACATCCCAAAAGCGATTGA 651
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGCTGGAAATAATATGCAACGAGT 711
Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATTCTTCAACAGCTAGTGACAT----- 744
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 745 -----AACACGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCA 783
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db 784 GCAAAATAATCTGAAATCTCCAGAGTCTTTTGAAGGAACTCTATCATTCACCTAGCGCC 843
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 844 CAACGTACAGTGAATCAGATGCGCTGTTTGGACCTGCTAAGATTATCAGTCGTACA 903
Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db 904 CCAATGGAGTTCGATTCGCGATGCGCACCATACCTTTATCTCTTACAGCAAGCTT 963
Qy 321 SerGluLeuGluGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTrp 340
Db 964 TCTGCTCTAGAGAAAGATGCGAAGATGGTGCCT----- 999
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 999 ----- 999
Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
```

```
Db 1000 -----ATCAGTGGAACTGGTTCTACAGTTCTTACAAATGCA 1035
Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db 1036 AAA----- 1038
Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1039 -----CCTAATGAAGTAGTGTCTAGTCTAGCAGCTCTTTCAGCAATCTCTTCT 1089
Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1090 -----TTAAACGACAAAGTAAGAGCTCTCTCAGCATCTGATGTTATATTTT 1137
Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db 1138 AATCCAAAA---GATATCGTTCAAGAAACGGCTACAGCTTATATTGTAAGACATGCT--- 1191
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuArgLeuAsnAspGluSerThrAsn 480
Db 1192 ---GATCATTTCCATTACATTCCAAAA-----TCAAATCAAATGGG 1230
Qy 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1231 CAACCGACTCTCCAAACAATAGTCTAGCAACCTCTCCCA---TCTCTCCAATCAAT 1287
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1288 CCAGAACTTCATCAGAAACATGAA----- 1314
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1315 -----GAAGATGATACGATTTGATGCTAATCGTATTATCGTGAA 1356
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db 1357 GATGAATCAGGTTTGTGTATGATGTCAGGAGACCAATCATATTATTCTTCAGAGGAC 1416
Qy 561 LeuSerAspLysGlu 565
Db 1417 TTGACAGAGAGCAA 1431
```

RESULT 12  
US-08-961-083-181  
; Sequence 181, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 181:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1342 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 US-09-661-083-181

## Alignment Scores:

Pred. No.: 147e-108 Length: 1342  
 Score: 1187.50 Matches: 257  
 Percent Similarity: 59.28% Conservative: 72  
 Best Local Similarity: 46.31% Mismatches: 117  
 Query Match: 28.51% Indels: 109  
 DB: 3 Gaps: 10

US-09-765-271-56 (1-796) x US-08-961-083-181 (1-1342)

```

QY      4 LeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyrIleAspGly 23
Db      2 CTAACAGCATCTTCGCGAGAAATAAGACAAATATCGTCTCTTATGTGGATGC 61
QY      24 LysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLysArgGluGly 43
Db      62 AGCAGTCAGTCAGAAAGTGAACCTTGACACAGACAGGTTAGCCAGAAAGAGA 121
QY      44 IleAsnAlaGluGlnIleValLysLeuThrAspGlnGlyTyrValThrSerHisGly 63
Db      122 ATTCAAGCTCAGCAATTTGTAATCAAAATTACAGATCAGGCTATGTAACGTCACACGGT 181
QY      64 AspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleLeuSerGluGluLeu 83
Db      182 GACCACATCATTAATAATGGAAGTTCCTTATGATGCGCTCTTTAGTGAAGAACTC 241
QY      84 LeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
Db      242 TTGATGAGGATCCNAATCACTTAAGACCGCTGATATGCAATGAAGTCAAGGGT 301
QY      104 GlyTyrValLysValAspGlyLysTyrTyrValTyrLysLysAspAlaAlaHisAla 123
Db      302 GGTATATCATCAAGTCGATGGAATAATATTATGCTACCTGAAAGATGCGCTCATGCT 361
QY      124 AspAsnValArgThrLysGluLysLysAsnArgGlnLysGlnGluHisSerGlnHisArg 143
Db      362 GATAATGTTGAACTAAAGATGAATCAATCGTCAAAACCAAGAACATGTCAAGATAAT 421
QY      144 GluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyr 163
Db      422 GAG-----AAGTTTAACTTAATGTTGCTGTAGCAAGTCTCAGGGACGATAT 469
QY      164 ThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleLeuGluAspThrGlyAspAla 183
Db      470 ACCACAAATGATGTTATGTCCTTAATCCAGCTGATATTATCGAAGATACGGGTAAATGCT 529
QY      184 TyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSer 203
Db      530 TATATGCTTCTCATGAGGTCACTATCACTACATTCCTCAAAACGCGATTATCTGCTAGT 589
QY      204 GluLeuAlaAlaGluAlaPheLeuSerGlyValArgGlyAsnLeuSerAsnSerArgThr 223
Db      590 GAATTACGACGCTAAGACATCTCGCTGGAAATAATATGCAACCGAGTCAGTTAAGC 649
QY      224 TyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSerAsn 243
Db      650 TATCTTCAACAGCTAGTGACAAAT----- 673
QY      244 ProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263
Db      674 -----AACACGCAATCTGTAGCAAAAGGATCAACTAGCAACGACCAATAA 721

```

```

QY      264 SerAsnAspIleAspSerSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHis 283
Db      722 TCTGAAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTAGCGCCCAAGCTTAC 781
QY      284 ValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThrAlaArgGly 303
Db      782 AGTGAATCAGATGCGCTGCTTTGACCTCTAAGATTATCAGTCGTACCAATATGGA 841
QY      304 ValAlaValProHisGlyAspHisTyrHisPheLeuProTyrSerGlnMetSerGluLeu 323
Db      842 GTTCGATTCGCGATGCGACCATTCACACTTTATTCCTTACAGCAAGCTTTCTGCCTTA 901
QY      324 GluGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHisTrpValProAsp 343
Db      902 GAAGAAAGATTGCCAAGATGGTGCT----- 928
QY      344 SerArgProGluGlnProSerProGluProThrProGluProSerProGlyProGlnPro 363
Db      928 ----- 928
QY      364 AlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuValArgLysVal 383
Db      929 -----ATCAGTGAAGCTGCTTCTACAGTTTCTACAAATGCAAAA----- 967
QY      384 GlyGluGlyTyrValPheGluGluLysGlyLysSerArgTyrValPheAlaLysAspLeu 403
Db      967 ----- 967
QY      404 ProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSerHis 423
Db      968 CCTAATGAAGTAGTGTCTAGCTAGGAGCTTTCAAGCAATCCTTCTCT----- 1018
QY      424 ThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLysAla 443
Db      1019 ---TTAAGCAGACAGTAAGGAGCTCTCTCAGCATCTGATGTTATATTTTAAATCCAAA 1075
QY      444 TyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAsp 463
Db      1076 ---GATATCCTTGAAGAAACGGCTACAGCTTATATTGTAAGACATGGT-----GATCAT 1126
QY      464 PheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsnLysGluLys 483
Db      1127 TTCAATTACATTCCAAAA-----TCAATCAATGGCAACCGACT 1168
QY      484 LeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuGlyLys 503
Db      1169 CTTCCAAACAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAATCCAGGAAC 1225
QY      504 ProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLys 523
Db      1226 TCATCGAGAAACATGAA----- 1243
QY      524 TyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAspGluGlyAsp 543
Db      1244 -----GAAGATGATACCGATTGTAGCTAATCGTATTATCGCTGAAGATGAATCA 1294
QY      544 AlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLys 558
Db      1295 GGTTTGTATGATGTCACGAGACCAATCAATATTATTTCTTCAAG 1339

```

## RESULT 13

US-09-536-784-181  
 ; Sequence 181, Application US/09536784  
 ; Patent No. 6573082  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland



```

Db 1244 -----GAAGATGGATACGGATTTCATCTAATCGTATTATTCGCTGAAGATCAATCA 1294
Qy 544 AlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLys 558
Db 1295 GGTTCATGAGTCACGGAGACCAACATCATTTATTCCTCAAG 1339

RESULT 14
US-08-961-527-355
; Sequence 355, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 355:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-355

Alignment Scores:
Pred. No.: 6.57e-89 Length: 973
Score: 988.00 Matches: 203
Percent Similarity: 79.79% Conservative: 26
Best Local Similarity: 70.73% Mismatches: 42
Query Match: 23.72% Indels: 17
DB: 4 Gaps: 4

US-09-765-271-56 (1-796) x US-08-961-527-355 (1-973)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
Db 158 TCCTATGAGCTTGGACGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 217
Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 218 TATATAGATGCTGATCAGGCTGGTCAAAAGGCGAGAAACCTTGACACAGATGAAGTCAGT 277
Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 278 AAGAGGAGGGGATCAACGGCGAACAATTTGTTATCAAGATTACGGATCAAGTTATGTC 337
Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db 338 ACCTCTCATGGAGACCATTTATCTACTATAATGGCAAGGTTCTTTATGATGCCATCATC 397

```

```

Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrIleLysLeuLysAspGluAspIleValAsn 99
Db 398 AGTGAAGAGCTCCTCATGAAAGATCCGAATATATCAGTTGAAGGATTCAGACATTTGTCAAT 457
Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 458 GAAATCAAGGGTGGTTATGTCATTAAGGTAAACCGTAAATACTATGTTTACCTTAAGGAT 517
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db 518 GCAGCTCATGCGGATAATATTCGGACAAAAGAGAGATTAACCTCAGAAGCAGGAAACGC 577
Qy 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
Db 578 AGTCATAATCATAACTCA-----AGACCAAGATAATGCTGTGTGCGACGAGAGCC 628
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 629 CAAGGACGTTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGAC 688
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 689 ACGGTGATGCTTATATCGTTCTCCACGGCGACCATTAACCATTAACATTCCTAAGAATGAG 748
Qy 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 749 TTATCAGCTAGCGAGTTAGCTGCTGCAGAACCCCTATTGGAAATGGG----- 793
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239
Db 794 -----AAGCAGGATCTCGCTCTTCAAGTCTCTGATTATAATGCA 835
Qy 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnThrSerAsnThrAsnSer 259
Db 836 AATCCAGCTCAACCAAGATTGTCAGAGAACCAACAATCTGCTGCTCACTCCAACTTATCAT 895
Qy 260 GlnAlaSerGlnSerAsnAspIleAspSerIleLeuLysGlnLeuTyrLysLeuProLeu 279
Db 896 CAA---AATCAAGGGGGAACATTTCAAGCCCTTTTACGTGAATTTGTATGCTAA-CCCTTA 951
Qy 280 SerGlnArgHisValGluSer 286
Db 952 TCAGAACGCCATGTGGGATCT 972

RESULT 15
US-08-961-527-258
; Sequence 258, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```



```
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 258:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-258

Alignment Scores:
Pred. No.: 7,446-67 Length: 1684
Score: 770.50 Matches: 157
Percent Similarity: 67.90% Conservative: 27
Best Local Similarity: 57.93% Mismatches: 58
Query Match: 18.50% Indels: 30
DB: 2
Gaps: 4

US-09-765-271-56 (1-796) x US-08-961-527-258 (1-1684)

QY 544 AlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAsp 563
Db 3 GCCTATGTAACCTCCACATATGACCATAGCCATGATTAAGAAAGATGTTGTCGAA 62

QY 564 LysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProSerPro 583
Db 63 GCTGAGAGCGGCCACC-CAGGCTTATGCTAAAGAGAAAGTTTGACCCCTCCTTCGACA 121

QY 584 AspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgVal 603
Db 122 GACCATCAGGATTCAGGAATCTCAGGCAAAAGGAGCAGAGCTATCTACACCCGCTG 181

QY 604 LysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGlu 623
Db 182 AAAGCAGCTAAGAGGTGCCACTTGATCGTATGCCTTACAACTTCAATATATCTAGAA 241

QY 624 ValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIleLysPheAla 643
Db 242 GTCAAAAACGGTGTATTAATCATCCTCATATGACCATACCAATACCAATCAAAATTGAG 301

QY 644 TrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAla 663
Db 302 TGGTTTGACGAGGCTTTATGAGGACCTAAGGGGTATCTCTTGAGGATCTTTGGCG 361

QY 664 ThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrpGly 683
Db 362 ACTGTCAGTACTATGTCGAACATCCAAACGAAACGTCGCCCATTCAGATAATGTTTGT 421

QY 684 AsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsnPhe 703
Db 422 AACGCTAGCAGCATGTTTCAAGAGAAACAAATATGTTCAAGCTGATACCAATCAACGGAA 481

QY 704 LysAlaAspGluGlu-----ProValGluGluThr----- 713
Db 482 AAACCAAGCGAGGAGAAACCTCAGACAGAAACCTTGAGGAGAAACCCCTCAGAGAGAG 541

QY 714 -----Pro 714
Db 542 AAACCCGAAAGCGAGAAACAGAGTCTCCAAAACCAACAGAGAGAACCAAGAAATCAACCA 601

QY 715 AlaGluProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAla 734
Db 602 GAGGAATCAGAGAAACCTCAGGTCGAGACTGAAAGGTTGAAGAAACAACTGAGAGAGCT 661

QY 735 GluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeu 754
Db 662 GAAGATTACTTGGAAAAATCCAGGATCCAAATATCAAGTCCATCCAAAGAGACTCTC 721

QY 755 AlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGlu 774
```

```
Db 722 ACAGGATTAAAAATAATTACTATTTCGACCCAGGACAAACAATACTATTATGCGAGAA 781
QY 775 AlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 782 GCTGAAAAACTATTGGCTTTATTAAAGGAGAGT 814

RESULT 16
US-08-743-637B-34/c
; Sequence 34, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: QUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-743-637B-34

Alignment Scores:
Pred. No.: 6,446-21 Length: 841
Score: 304.50 Matches: 74
Percent Similarity: 59.24% Conservative: 19
Best Local Similarity: 47.13% Mismatches: 29
Query Match: 7.31% Indels: 36
DB: 2
Gaps: 3

US-09-765-271-56 (1-796) x US-08-743-637B-34 (1-841)

QY 663 AlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
Db 836 GCGACTGTCACTACTATGTCGAACATCCAAACGAGCGTCG-CATTGAGATAATGTTT 778

QY 683 GlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLysAsn 702
```

```

; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; US-08-526-840B-34

Alignment Scores:
Pred. No.: 6,44e-21 Length: 841
Score: 304.50 Matches: 74
Percent Similarity: 59.24% Conservative: 19
Best Local Similarity: 47.13% Mismatches: 29
Query Match: 7.31% Indels: 36
DB: 3 Gaps: 3

US-09-765-271-56 (1-796) x US-08-526-840B-34 (1-841)

Qy 663 AlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
Db 836 GCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCG-CATTTCAGATAATGGTTT 778
Qy 683 GlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsn 702
Db 777 GGTAACCTAGCCACCATGTTCAAAAGAAACAAATAATGTCAGCTGATACCAATCAAACG 718
Qy 703 PhelysAlaAspGluGlu-----ProValGluGluThr----- 713
Db 717 GAAACCAACACGAGGAGAAACCTCAGACAGAAACCTCAGGAGAGAAACCCCTCGAGAA 658
Qy 713 ----- 713
Db 657 GAGAAACCGCAAGCGAGAAACCGAGTCTCCAAACCAACAGAGGAAACGAGAAAGAA 598
Qy 714 -----ProAlaGluProGluValProGluValGluThrGluLysValGlu 728
Db 597 TCACCAGAGAAATCACCAGAGGAAATCAGAGAACCTCAGGTCGAGACTGAAAGGTTAAGA 538
Qy 729 AlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 537 ---AACTGAGAGGCTGAGATTTACTTGGAAATAATCCAGAAATCCAATTATCAAGTCC 481
Qy 749 AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsn 768
Db 480 AATGCCAAAGAGACTCTCAGAGGATTAAATAATTTACTATTGTCACCCAGGACAAAC 421
Qy 769 AsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 420 AATACTATTATGCGAGAGCTGAAATACTATTGGCTTTATTAAGGAGAGT 370

RESULT 17
US-08-526-840B-34/c
; Sequence 34, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: GUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; US-08-526-840B-34

Alignment Scores:
Pred. No.: 6,44e-21 Length: 841
Score: 304.50 Matches: 74
Percent Similarity: 59.24% Conservative: 19
Best Local Similarity: 47.13% Mismatches: 29
Query Match: 7.31% Indels: 36
DB: 3 Gaps: 3

US-09-765-271-56 (1-796) x US-08-526-840B-34 (1-841)

Qy 663 AlaThrIleLysTyrValGluHisProAspGluArgProHisSerAsnAspGlyTrp 682
Db 836 GCGACTGTCAAGTACTATGTGCAACATCCAAACGACGTCG-CATTTCAGATAATGGTTT 778
Qy 683 GlyAsnAlaSerGluHisValLeuGlyLysAspHisSerGluAspProAsnLysAsn 702
Db 777 GGTAACCTAGCCACCATGTTCAAAAGAAACAAATAATGTCAGCTGATACCAATCAAACG 718
Qy 703 PhelysAlaAspGluGlu-----ProValGluGluThr----- 713
Db 717 GAAACCAACACGAGGAGAAACCTCAGACAGAAACCTCAGGAGAGAAACCCCTCGAGAA 658
Qy 713 ----- 713
Db 657 GAGAAACCGCAAGCGAGAAACCGAGTCTCCAAACCAACAGAGGAAACGAGAAAGAA 598
Qy 714 -----ProAlaGluProGluValProGluValGluThrGluLysValGlu 728
Db 597 TCACCAGAGAAATCACCAGAGGAAATCAGAGAACCTCAGGTCGAGACTGAAAGGTTAAGA 538
Qy 729 AlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 537 ---AACTGAGAGGCTGAGATTTACTTGGAAATAATCCAGAAATCCAATTATCAAGTCC 481
Qy 749 AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsn 768
Db 480 AATGCCAAAGAGACTCTCAGAGGATTAAATAATTTACTATTGTCACCCAGGACAAAC 421
Qy 769 AsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 420 AATACTATTATGCGAGAGCTGAAATACTATTGGCTTTATTAAGGAGAGT 370

RESULT 18
US-08-956-171E-63
; Sequence 63, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E

```

```

; FILLING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILLING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILLING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 63:
US-09-765-171E-63

```

```

Alignment Scores:
Pred. No.: 6,39e-06 Length: 8155
Score: 170.00 Matches: 169
Percent Similarity: 35.36% Conservative: 139
Best Local Similarity: 19.40% Mismatches: 355
Query Match: 4.08% Indels: 208
DB: 4 Gaps: 38

```

```

US-09-765-271-56 (1-796) x US-08-956-171E-63 (1-8155)
QY 8 GlnAlaArgThrValLysGluAsnArgValSerTyrIleAspGlyLeuGlnAlaThr 27
DB 4948 CAAGCTGCCAACAATAAGATCGATGCAAAAGCGAAATCGCTCAAAAAGCAAGTGA 5007
QY 28 GlnLysThr-----GluAsnLeuThrProAspGluValSerLysArg 41
DB 5008 COTAAACAGCAATTAAGCAATGAATGATCGACTACTGAGAACACACAGCAGCGAAA 5067
QY 42 GluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSer 61
DB 5068 GACAAAGTG-----GATCAAGCAGTAGTACTGCA 5097
QY 62 HisGlyAsp-----HisTyrHisTyrTyrAsnGlyLysValProTyr 75
DB 5098 AACGCTGATATAGATAATGCTGCAGCAACCAATGATGTAATGCAAAAACACTACAAAT 5157
QY 76 AspAlaIleIleSerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGlu 95
DB 5158 GAAGCTACAAUGCA-----GCCATTACCTGATGCAAAATGTTAAACCCAGCAGCAAA 5211
QY 96 AspIleValAsnGluValLysGlyTyrValIleLysValAspGlyLysTyrTyrVal 115
DB 5212 CAAGCAATTGCAGATAAAGTCAAGCTCAAGAAACAGCAATTTGATGGA-----5259
QY 116 TyrLeuLysAspAlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGln 135
DB 5260 -----AATAACGGCTCAACCACTGAAGAAACAGCAGTGT 5295
QY 136 LysGlnGluHisSerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAla-----153
DB 5296 AAACACACAGTTCAAACTGAAACCAACACAGCTGATGCGCAATAGATGACACATACA 5355
QY 154 -----ValAlaLeuAlaArg-----SerGlnGly 161
DB 5356 AATCGCGAAGTTGAAGCGCTTAAAGAGCAGCAATTTGCTTAAATTAAGCGGATTCAGCCA 5415
QY 162 ArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGly 181
DB 5416 GCAACACCACTAAAGAT-----AATGCGAAAGACCAATTCCTACGAAGCG 5463

```

```

QY 182 AspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSer 201
DB 5464 AATGAACGTAAACACAGCA-----ATCGCTCAACGCAACGACATTACT 5505
QY 202 AlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSer 221
DB 5506 GCTGAAGAAATTCAGCGGCTTAATGCG-----GACGTAGATAATGCT 5547
QY 222 ArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSerVal 241
DB 5548 GTGACA-----CAAGCAATAGCAACATTGAAGCTGCTAATAGTCAAAATGATGTA 5598
QY 242 SerAsnProGlyThrThrAsnThrAsnThrSerAsn-----AsnSerAsnThrAsnSer 259
DB 5599 GACCAAGCGAAACGACACAGGTGAAATAGTATTGATCAAGTAACACCAACAGTTAATAAA 5658
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
DB 5659 AAAGCAACTGCACGTAAATGAATCAGCAATTTTAAATAAC-----AAATTGCAAGAG 5712
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
DB 5713 ATTCAAGCTACGCGACGATGCAACAGATGAAGAAACACAGCGCTGCTGTAAGCAAT 5772
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
DB 5773 ACTGAAATGCTAAAGCA-----AATCAA 5796
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339
DB 5797 GCCATTTTCAGCAGCAACTACTAAGCAAGTGTGATGAAGCTAAAGCAAAATGCAAGACA 5856
QY 340 TrpValProAspSerArgProGlnProSerProGlnProThrProGluProSerPro 359
DB 5857 CGGATTAATCGGTAAACCAAAAGTTGTGAAGAAACAGCGGCTAAAGATCAAAATGAT 5916
QY 360 GlyProGlnProAla-----ProAsnLeuLysIleAspSer 371
DB 5917 CAATTACAAACGCAACCAAAATGTTATCAATATGATCAGAACGCTACACACAGAGAA 5976
QY 372 AsnSerSerLeuValSerGlnLeuValArgLysValGlyGluGly-----386
DB 5977 AAAGAACGAGCTATTCAACAATTTAGCAACAGCAGTTACAGACGCGCAAAATAATATTACA 6036
QY 387 TyrValPheGluGluLysGlyIleSerArgTyrValPheAlaLysAspLeuProSerGlu 406
DB 6037 GCTCAACTGATGATGATGTTGTTAGTACAG-----GCGAAAGACGCTGGAAGAAT 6087
QY 407 ThrValLysAsnLeuGluSer-----LysLeuSerLysGlnGluSerValSer 422
DB 6088 TCAATTCAAAAGCACGCAACCAACAGCAACAGCGGTTAAATCAAAATGCTTAAATGATGTTGAT 6147
QY 423 HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAspLys 442
DB 6148 CAAGCTGTGACAACTCAAAATCAAGCAAT-----GATAAT 6183
QY 443 AlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu-----Phe***AsnLysGlyArg 460
DB 6184 ACAACTGCTGTCTACAACTGAAGAGAAAAATGCAAGCAAAAGATTTAGTTTAAAGCTAAA 6243
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
DB 6244 GAAAAAGCGTATCAAGATATC-----TTAAATGCAACAAACCAACTAAT 6285
QY 481 -----LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProlle 495
DB 6286 GATGTACGCAAAATTAAGATCAAGAGTTGCTGTATTTCAAGGTATTACTGCAGATACA 6345
QY 496 ThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluVal 515
DB 6346 ACAATTAAAGAT---GTTGCGAAAGATGAATTAATGACCAAAAGCAACCAAAAGCG 6402
QY 516 ArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHis 535

```

```

Db 6403 CTTATTGCACAACTGCAGATCGGACTACT.....:---GAAGAAAA 6441
Qy 536 AspilelleSerAspGluGlyAspAlaThrProHisMetGlyHisSerHisTrp 555
Db 6442 GAACAGCAATCAACAGTAGCCCAAAATTAACA-----CAGGTAATCAAAATATT 6495
Qy 556 IleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaThrLysGlu 575
Db 6496 GAAATGACACAGTCAATCGATGATTAACACTGCAAGGATAATGCAATT----- 6546
Qy 576 LysGlyLleLeuProSerProAspAlaSerValLysAlaAsnProThrGlyAspSer 595
Db 6547 CAACGAATTCACCAATTCACAGCATCAACAGATGTTAAACGAATGCAAGCGCAATTG 6606
Qy 596 AlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgLleProLeuValArgLeuPro 615
Db 6607 CTAACGAAATGCAAAATAAAATA-----ACTGAAATACTT 6642
Qy 616 TyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlePro----- 632
Db 6643 AATAATAATGAGACTACTAATGAAGAAAA---GGTAACGATATTGACCGATTAGACGA 6699
Qy 633 ---HisLysAspHisTyrHisAsnLleLysPheAlaTrpPhe-----AspAspHisThr 649
Db 6700 GCATATGAAGAAGGTTTAAATAATTAATTAATGCAGCACTACTACAGGTGATTAATACT 6759
Qy 650 TyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrLleLysTyrVal 669
Db 6760 GCTAAGATACAGCAGTACAAAAGTTCAACAACTTCATGCA----- 6801
Qy 670 GluHisProAspGluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisVal 689
Db 6802 ---AATCCTGTTAAGAAACCA----- 6819
Qy 690 LeuGlyLysLysAspHisSerGluAspProAsnLysAsnPheLysAlaAspGlu----- 708
Db 6820 GCAGGTAAAAAAGAAATTAGATCAAA-----GCTGCAGCTGATAGAAAAACA 6864
Qy 709 ProValGluThrPro-----AlaGluProGluValProGlnValGluThrGluLys 726
Db 6865 CAATAGAACAAACACCAATATGTCATCAACAAAGAAATTAATGATGCAAAACAGAA--- 6921
Qy 727 ValGluAlaGlnLeuLysGluAla-----GluVal 736
Db 6922 GTTGATCTGAATTAATAATCAAGCGAAACAAATGTCATCAATCATCAACAAATGATAT 6981
Qy 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly 756
Db 6982 GTTGATAATGCAGTTAAAGAAAGAAAGCTTAAATAATTAATGCAGTTAAACATTTAGTGAG 7041
Qy 757 LeuArgAsnAsnLeuThrLeuGlnIleMetAsp----- 767
Db 7042 TACAAAAGATGCTTTAGCTAAATTTGAAGATGTCATATTAATGCTAAAGTAAACGAGCG 7101
Qy 768 -----AsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeu 781
Db 7102 GATACTCTAACGCATCGACTTCAAGTGAATTTGCTGACGGAACAAACAACTTGCTGAA 7161
Qy 782 LeuLysGlySerAsnProSerSerValSerLys 792
Db 7162 TTAANAACAACTCGCGATCAAAATGTTAATCAA 7194

```

## RESULT 19

US-08-956-171E-310  
 ; Sequence 310, Application US/08956171E  
 ; Patent No. 6593114

## GENERAL INFORMATION:

APPLICANT: Charles Kunsch  
 ; Gil H. Choi  
 ; Patrick S. Dillon  
 ; Craig A. Rosen  
 ; Steven C. Barash

Michael R. Fannon  
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 NUMBER OF SEQUENCES: 5256  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2

## SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E  
 FILING DATE: 20-Oct-1997

## CLASSIFICATION: &lt;Unknown&gt;

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861  
 FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986  
 FILING DATE: January 3, 1997

## ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman  
 REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224  
 TELEFAX: (301) 309-8439

## INFORMATION FOR SEQ ID NO: 310:

## SEQUENCE CHARACTERISTICS:

LENGTH: 12173 base pairs  
 TYPE: nucleic acid

STRANDEDNESS: double  
 TOPOLOGY: linear

## SEQUENCE DESCRIPTION: SEQ ID NO: 310:

US-08-956-171E-310

## Alignment Scores:

Pred. No.:	1-54e-05	Length:	12173
Score:	169.00	Matches:	179
Percent Similarity:	34.99%	Conservative:	138
Best Local Similarity:	19.76%	Mismatches:	343
Query Match:	4.06%	Indels:	246
DB:	4	Gaps:	46

US-09-765-271-56 (1-796) x US-08-956-171E-310 (1-12173)

```

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeu---ThrProAspGluValSer 39
Db 9540 ATTACTTCTCAACATCAAGCAGCAGCAAGAAATACAAATACTTCAGATAAAATCTCG 9599
QY 40 LysArgGluGlyLleAsnAlaGluGlnIle---ValleLysIleThrAspGlnGlyTyr 58
Db 9600 GAAATCAAAATAAATAATGCAATCAACTACACTCACCACCTAAGGATACAAATCAACACAA 9659
QY 59 ValThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIle 78
Db 9660 CCGTCTACGCAACCA-----GCACACACTGCGGAAACACTATCTCTGCA--- 9701
QY 79 IleSerGluGluLeuLeu-----MetLysAspProAsnTyrHisLeuLysAspGlu 95
Db 9702 ---GCGGATGAATCACTTAAGATGCAATTAAGATCTCTGCATTCAGAAAATAAGAACAT 9758
QY 96 AspilelValAsnGluValLysGlyGlyTyrValIle-----LysValAspGlyLys 112
Db 9759 GATATAGGTCCCAAGACAAACAGTCAATTTCCAGTTATTATAGATAAAACAAATGAACGCAG 9818
QY 113 TyrTyrValTyr-----LeuLysAspAlaAlaHisAlaAsnValArgThrLysGlu 130
Db 9819 TACTATCACTTTTTCAGCATCAAGATCCAGCAGATGTGTATTACACTAAAGAAAGCA 9878

```

```
Qy 131 GluIle-----Asn 133
|||
Db 9879 GAAGTTGAATTAGACATCAATACTGCTTCAACATGGGAAGATTGGAAGTCTATGAAAAC 9938
|||
Qy 134 ArgGlnLys-----GlnGluHisSerGln 141
|||
Db 9939 AATCAAAAATGGCCAGTCAGACTTGTATCATATAGTCCTGTACCAGAAGACCATGCGCTAT 9998
|||
Qy 142 HisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGly 161
|||
Db 9999 ATTCCGA-----TTCCAGATTTCAGATGCCACACAAGAATTG--AAAATGTTTCT 10046
|||
Qy 162 ArgTyrThrThrAspAspGlyTyr-----169
|||
Db 10047 TCGACTCAAAATGATGGAGGAGAGAAACAAATATGATGTTATCTAAATTAGTATTGCT 10106
|||
Qy 170 -----IlePheAsnAlaSerAspIleIleGlu---AspThrGlyAspAlaTyrIleVal 186
|||
Db 10107 AAACCTATTATACGATCCTTCACTTGTAAAATCAGATACAAATGATGTCAGTAGTAACG 10166
|||
Qy 187 ProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAla 206
|||
Db 10167 -----AATGATCAATCAAGTTTCAGTCGCAAGTAAATCAAAACAACAGATACATCTAAT 10220
|||
Qy 207 AlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArg 226
|||
Db 10221 CAAAATATATCAACGATCAACATCTAATATCAACCGCAGGCAACGACCAATATGAGT 10280
|||
Qy 227 GlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSerAsnProGly---245
|||
Db 10281 CAACCTGCACACCAAAATCGTCAAGAAATGCAGATCAAGCGTCAAGCAACAGCTCAT 10340
|||
Qy 246 ThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsn 265
|||
Db 10341 GAAACAAATTTCTAATGGTAATCTAATCAAGTAAACGAAATGAGTCAAGTAATCAGTCGGAT 10400
|||
Qy 266 AspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGlu 285
|||
Db 10401 -----GTTAATCAACAGTATCCACCA 10421
|||
Qy 286 SerAspGlyLeuValPheAsp-----ProAlaGlnIleThrSerArgThrAla 301
|||
Db 10422 GCAGATGAATCACTACAGATGCAATTTAAACCCGGCTATCATCGATAAAGAACATACA 10481
|||
Qy 302 ArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSer 321
|||
Db 10482 -----GCTGATAATTCGGCAGCAATTTGATTTTCAAATGAAAAT 10520
|||
Qy 322 GluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrpVal 341
|||
Db 10521 GATAAAGGTGAAGA-----CAGTTCTATCATATTGCT 10553
|||
Qy 342 ProAspSerArgProGluGlnProSerProGlnProThrProGluProSer-----358
|||
Db 10554 -----AGTACTGTTGAACCCAGCAACTGTCAAT 10580
|||
Qy 359 -----ProGlyProGlnProAlaProAsnLeuLys-----368
|||
Db 10581 TTTACAAAAACAGGACCAATAATGAAATGAGTTTAAAGACAGCTTCAACATGGGAAGAAA 10640
|||
Qy 369 -----IleAspSerAsnSerSerLeuValSerGlnLeuValArgLysValGlyGlu 385
|||
Db 10641 TTTGAAGTTTGAAGTGACAAAAGTTACCAGTCGAATTAGTATCATATGATTTCTGAT 10700
|||
Qy 386 -----GlyTyrVal---PheGluGluLysGlyIleSerArgTyrValPheAlaLys 401
|||
Db 10701 AAAGATTATGCTATATTCGTTTCCAGTATCTAATGTCGACAGAGAAGCTT-----10751
|||
Qy 402 AspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerVal 421
|||
Db 10752 -----AAAATTGTGTCATCTATGATGATGTTGTCAGNACATCCATGAGACTAT 10799
|||
```

```
Qy 422 SerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyrAsp 441
|||
Db 10800 GATTATACGCTAATGCTTTTGACACGCTATTATTAATCAACCCAGACGACTATGTCGAT 10859
|||
Qy 442 ---LysAlaTyrAsn-----LeuLeuThrGluAlaHisLysAlaLeuPhe**Asn 457
|||
Db 10860 GAAGAAACATACAAATTTACAAAATTTATAGTCCGTATCACAAGCT-----10907
|||
Qy 458 LysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGlu 477
|||
Db 10908 AAAACGTTAGAAACACACAAAGTTTATGAATAGAAAAATTAACAAGAGAAATTCGCCAGAAAA 10967
|||
Qy 478 ---SerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla-----493
|||
Db 10968 TATAAGCGCGAATATAAAAAAGAAATTTAGATCAAACTAGAGTAGAGTTCAGTCAAGTT 11027
|||
Qy 494 -----ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThr 511
|||
Db 11028 AAATCAGCAGTGACGGAATTTGAAAATGTTACACTACAAATGATCAATTAACAGATTTA 11087
|||
Qy 512 GluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSer-----AspGly 529
|||
Db 11088 CAAGAAGCGCATTTTGTGTTTTTGAAGTGAAGAAATAGTGAAGTTCAGTTATGACGCGC 11147
|||
Qy 530 TyrIlePheAspGluHisAspIleIleSerAsp-----GluGlyAspAlaTyrValThr 547
|||
Db 11148 TTTGTT-----GAACATCCATCTATACACCACTTTAAATGTTCAAAAATATATAGTG 11201
|||
Qy 548 ---ProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLys 566
|||
Db 11202 ATGAAAAACAAAGGATGACAGTTACTGG-----AAAGATTAAATTTGTAGAAGTTAAACGT 11255
|||
Qy 567 ValAlaAlaGlnAlaTyrThrLysGluLysGly-----IleLeuProProSer 582
|||
Db 11256 GTCACTACTGTTCTTAAAGATCCTAAAAATAATTTCTAGAACGCTGATTTTCCATATATA 11315
|||
Qy 583 ProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyrAsnArg 602
|||
Db 11316 CTGAC-----AAACGAGTTTACACATGCG 11339
|||
Qy 603 ValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrVal 622
|||
Db 11340 ATT-----GTTAAAGTCGTTGTGCAACATTTGTTATGAAGTCAATATCATGTC 11390
|||
Qy 623 GluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle-----640
|||
Db 11391 AGAATTATAAATCAGGATATCAATACAAAAGATGATGATACATACAAAATAAACAGAGT 11450
|||
Qy 641 -----LysPheAlaTrpPheAsp-----Asp 647
|||
Db 11451 GAACCGCTAAATGTACAAACAGGACAGAGAGGTAAGTTGCTGATACAGATGTAGCTGAA 11510
|||
Qy 648 HisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyr 667
|||
Db 11511 AATACGACGACTGCAACAAATCCTAAAGATGCGTCTGATAAAGCAGATGTGATAGAA---11567
|||
Qy 668 TyrValGluHisProAspGluArgProHisSerAsn-----AspGlyTrpGlyAsn 684
|||
Db 11568 -----CCAGAGTCTGACGTTGTTAAAGATCTGCTGATATAAT 11603
|||
Qy 685 AlaSerGluHisValLeuLeuGlyLysAspHisSerGluAspProAsnLysAsnPhelLys 704
|||
Db 11604 ATTGATAAAGATGTCACACATGATGTTGATCTATTTATCCGATATGTCGGAATAATATCAC 11663
|||
Qy 705 AlaAspGluGluProValGluGluThrProAlaGluProGluValProGlnValGluThr 724
|||
Db 11664 TTCGAT-----11669
|||
Qy 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744
|||
Db 11670 -----AAATATGATTTAAAGAAATGGATCTCAAAATTCGCAAGATCTGATAGA 11720
|||
Qy 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeu---763
|||
```

```

Db 11721 AATGTGGATAAAGATGCCGATAATAGCGTT---GATGTGTCATCTAATGTGCTGATACTGAT 11777
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 764 -----GlnIleMetAspAsnAsn 769
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 11778 AAAGACTCTAATAAATAAAGACAAAGTCATACAGCTGAATCATATATTGCCGATAAATAAT 11837
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 770 SerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly---SerAsnProSer 788
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 11838 AATCATACTGGAAGACGACAAAGCTTGACGTAGTGAACAAAATTTATATATATACAGAC 11897
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 789 SerValSerLysGluLys 794
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 11898 AAAGTTACTGACAAAAA 11915

RESULT 20
US-09-134-001C-2243
; Sequence 2243, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2243
; LENGTH: 11091
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2243

Alignment Scores:
Pred. No.: 5,24e-05 Length: 11091
Score: 163.00 Matches: 184
Percent Similarity: 32.89% Conservative: 141
Best Local Similarity: 18.62% Mismatches: 347
Query Match: 3.91% Indels: 316
DB: 4 Gaps: 43

US-09-765-271-56 (1-796) x US-09-134-001C-2243 (1-11091)
Qy 15 AsnAsnArgValSerTyrIle-----AspGlyLysGlnAlaThrGlnLysThrGluAsn 32
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3895 AATAATCAATTCATGGTATTTGTGAGTGTGTAGTGTGATGACAAAGCATAAATGCA----- 3945
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 33 LeuThrProAspGluValSerLysArgGluGlyIleAsnAlaGluGlnIleValIleLys 52
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 3946 ATTACACCTGATACATCAATTAAGAAGATGCTAAATAAT-----GATATTGATATATAA 3999
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 53 IleThrAsp-----GlnGlyTyrValThrSerHisGlyAspHisTyrHis 67
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4000 GCAGCTGATAAGAAAATAAAAAATTCAAAGATAAATGATGCTACAGATGAAGAAATTCAA 4059
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 68 TyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGluLeuLeuMetLysAsp 87
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4060 GAAGCGAATTCGTAATAATTGAA-----GAAGCTAAGATTGAAGCAAAAGAT 4104
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 88 ProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIle 107
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4105 AATATTCAAGCGAATAGTACTAGAGATCAAGTAAATGAAGCGAAAACTAATGGAATAAAT 4164
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 108 LysValAsp----- 110
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4165 AAAATAGAAAAATAACACCAGCAACTACTGTGAAATCTGACCTAGACAGCAGGTACAG 4224
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 111 -----GlyLysTyrTyrValTyrLeuLysAspAlaIleHisAlaAspAsnValArg 127
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

Db 4225 AATAAGCAAAATGATGAACAGATTAAATCATATATTCAAAACACGCCTGATGCAACTAATGAAGAA 4284
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 128 ThrLysGluGluIleAsnArg-----GlnLysGlnGluHis 139
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4285 AAACAAGAGGCAATAAATAGTAGTAACTGCTGAATTAGCAAGAGTTCAAGCACAATAAAT 4344
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 140 SerGlnHisArgGluGlyGly-----ThrProArgAsnAspGlyAlaValAlaLeuAla 157
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4345 GCAGAACATACACCAAGGTTGTCAAACATATCAAGACGACGCGATAAATCTTTATCT 4404
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 158 ArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIle 177
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4405 CGA-----ATTAATGCAAGATTGTTGAG 4428
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 178 GluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLys 197
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4429 AAAGAG----- 4434
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 198 AsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsn 217
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4435 -----TCTGCAAGAAATGCA 4449
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 218 LeuSerAsnSerArgThrTyrArgArgGln-----AsnSerAspAsnThrSer--- 233
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4450 ATCGAACAAAGGCAACACACAAACGCAATTTATTATATATATATATGCTACAGAT 4509
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 234 -----ArgThrAsnTyrValProSerValSerAsnProGlyThrThrAsn 248
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4510 GAAGAAAAGAGGTGCGCAACAATTTAGTTATCGCTACAAAACAAAAATCATTAGATAAT 4569
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 249 ThrAsn-----ThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAsp 266
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4570 ATTAACCTCCTTATCTTCAATAATGATGTTGAAAATGCTAAAGTAGCAGGAATAAATGAA 4629
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 267 IleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSer 286
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4630 ATAGCTAACGTTTAA----- 4644
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 287 AspGlyLeuValPheAspProAla---GlnIleThrSerArgThrAlaArgGlyValAla 305
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4645 -----CCAGCAACGCTGTTAAGTCAAAAGCAAAAGAAAGATATTGAT 4686
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 306 ValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGluLeuGluGlu 325
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4687 CAAAAACTCGCCCAACAGATTAAATCAAAATCAAAACGATCAAACTGCTACAACCTGAGAA 4746
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 326 ArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis----- 339
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4747 AAAGAAGCGCTATTCAATTGGCAATCAAAATCAAAATCAAAAGCAAGCAAGCAATTCAA 4806
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 339 ----- 339
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4807 AATGAACATAGTAACAATGGTGTGCAACAAGCTAAATCTAACGGCATTCTATGAATTTGAA 4866
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 340 ---TipValProAspSerArgProGluGlnProSerProGlnProThrPro----- 355
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4867 TTAGTTATCCAGATCGGCACAAAATACTGATCTAAACAAAGTAGTATCGATAATAATAT 4926
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 356 ---GluProSerProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSer 374
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4927 AATGCAAAAGTAATACTATCAACTACACCATGACAGATGCAAGATGAAGAAAGCAAAAG 4986
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 375 LeuValSerGlnLeu-----ValArgLysValGlyGlyTyrValPheGluGluLys 392
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 4987 GCATTAGATAAATAAATAAGTAGTAAAGATGCAAGATACCAACAAAGTTGATCAAGCGCAA 5046
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 393 GlyIleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysLeuGlu 412
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 5047 ACAACCAACCAAGTATCTGATCGAAAACTGAGGCTATAGATACGATACTAATATTCAA 5106
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 413 SerLysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGlu----- 430
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 5107 GCAATGTTGCAAAAAAACCATCCGCTCGAGTGGAAATAGATTCAAGTTGAGGATTTA 5166
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```





```

Db      7501  -----|||:|:|:|-----ATCAATGAA-----ATTGTT 7515
Qy      86  LysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGluValLys----- 102
Db      7516  AAAGAACAGGGTCAAAATGTGACTAATGATGATATTAATTAATGCGGTTCAAGTGCCAAAT 7575
Qy      103  -----GlyClyTyr 105
Db      7576  AAAAATAGAGTTGCGATTAAACAGGAACGCTCTTCCAAACAATTTAGCTGGTGTAGT 7635
Qy      106  ValIleLysValAspGlyLysTyrValTyrLeuLysAspAlaAlaHis----- 122
Db      7636  ACATCACATATT-----CCAGTAGTTATTATTACAGTGATGGAAGTCTCGAAGAGCT 7689
Qy      123  AlaAspAsnValArgThrLys-----GluGluIleAsn-----ArgGlnLysGln 137
Db      7690  ACTGAGACTGTTAGAACTAAAGTTAAATAAACCCGAATTAATCAATGCTGCTGCTGACTA 7749
Qy      138  GluHisSerGlnHisArgGluGlyThrPro-----ArgAsn-----AspGlyAla 153
Db      7750  GATCAAGAAATTAGTAAAGAGAACAAACACCATCAAGTATCAGAACTTTGATCAAGCT 7809
Qy      154  ValAlaLeuAlaArgSerGlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAla 173
Db      7810  ATGAATCGTGTCTCAATCACAAATTAATACAGCTAAAGTGAT-----GCT 7854
Qy      174  SerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHis 193
Db      7855  GACCAAGTTATA-----GGCACAGAAATTTGCAACACCTCAACAAGTAAAT---TCA 7902
Qy      194  TyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSer 213
Db      7903  GCTTTATCTAAAGTTCAAGCGGCACAAATAAATAAATAGTAAAGCAATTTATCAA 7962
Qy      214  GlyArgGlyAsnLeuSerAsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSer 233
Db      7963  AACAAAGCTCGAT-----AATAGTCAACTTGTGAGACAAAGAACAAATACACAAATCG 8016
Qy      234  ArgThrAsnTrpValProSerValSerAsnProGlyThrThrAsnThrAsnThrSerAsn 253
Db      8017  ATTCAA-----CCAGCGCTTCAACTGATGATGATGATGATGATGATGATGATGATGATGAT 8067
Qy      254  AsnSerAsnThrAsnSerGlnAlaSerGlnSer----- 264
Db      8068  TACAAAATAAACCAGCAGCTGAACACAGCAATACAAATGCAATACGTTATATAAT 8127
Qy      265  -----AsnAspIleAspSerLeuLeuLysGlnLeuTyr 275
Db      8128  AATGGAGATGCAACATCCCAACAAATTAATGATGCTAAAAACACAGTTGAACAGGCACAG 8187
Qy      276  Lys-----LeuProLeuSerGlnArgHisValGluSer----- 286
Db      8188  AGAGATTATGTTGAAGCTAAAGCAACTTACGCTGCTGATAGTACACAGTTACAAAGCGCT 8247
Qy      287  -----AspGlyLeuValPheAsp-----ProAlaGlnIle----- 296
Db      8248  TATGATACGTTAAATAGAGATGTTTAAACAAATGATAAAAGCCAGCATCTGTAAGACGC 8307
Qy      297  -----ThrSer 298
Db      8308  TATAATGAAGCCATTTCAAATATTAGAAAAAGAAATTAGATACAGCTAAAGCGGATGCAAGT 8367
Qy      299  ArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSer 318
Db      8368  AGTACTTTGGAAACACC-----AATCCT 8391
Qy      319  GlnMetSerGluLeuGluGluAlaArgIleLeuProLeuArgTyrArgSerAsn 338
Db      8392  TCGTTTGAACAAGTTAGAGACGCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8451
Qy      339  HisTrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer 358

```

```

Db      8452  CAAGCAATTGCTTTACTTCAACCAAAAGAAATAAATTCAGAA---CTTGTACAAGCTAAA 8508
Qy      359  ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGln 378
Db      8509  AAACGTTTTACAAGACGCTGTAAAT---GACATACCTCAACACACAGGTATGACACAACA 8565
Qy      379  LeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrVal 398
Db      8566  ACAATTATTAATTAATTAATGACAAACACAGCTGAAGCTGAAGCACTTTACATCTGCACAA 8625
Qy      399  PheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGln 418
Db      8626  AGAGTGATTGATTAATCGGGATGCTCAAACTCAAGAAATTAATCTTGAAAAATCTAAAGTA 8685
Qy      419  GluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGlu 438
Db      8686  GAGCAAGCAATGCAAGCTTTAACTAATGCTTAAAGTAATCTGAGA---GCTGATAAGAAAT 8742
Qy      439  PheTyrAspLysAlaTyrAsnLeuThr----- 448
Db      8743  GAGTTACAGACTGCTATATAACAAATTAATTGAGAACGCTATCTACCAATGGTAAAAAACCG 8802
Qy      449  -----GluAlaHisLysAlaLeuPhe***AsnLys----- 458
Db      8803  GCGAGTATACGTCAATACGAAACACAGCCCAAGCCAGATACAAATCAAAATTAATTAATGATGCT 8862
Qy      459  -----GlyArgAsnSerAsp----- 465
Db      8863  AAAAATGAAGCGGAGCGAATTTTAGTGAATGATATAATCCCAAGTATCACAAGTAACTCAA 8922
Qy      466  AlaLeuAsp-----LysLeuLeuGluArgLeu----- 474
Db      8923  GCATTGAACAAAATCAAAGCTATTTCACCAAAAATTAACAGAACTATCAACATGCTTCAA 8982
Qy      475  -----AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAla 490
Db      8983  AACAAAGAAATTAATACAGATTAGTCAATGCTTAAACACAGACTTGAAAATGCGATAAAT 9042
Qy      491  PheLeuAlaProIleThrHis-----ProGluArgLeuGlyLysProAsnSerGln 507
Db      9043  GATACAGATCCA---ACACACGGTATGACTCAAGAAACAATTAATAATTACACAGCTAAA 9099
Qy      508  IleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrSer 527
Db      9100  AACGGAAGAGCTCAAAATGAATACAAAAGCG----- 9132
Qy      528  AspGlyTyrIlePheAspGluHisAspIleIleSerAspGluGlyAspAlaTyrValThr 547
Db      9133  -----ACATGATTATTAATTAATGAGATGCT----- 9159
Qy      548  ProHisMetGlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysVal 567
Db      9160  -----ACTGCTCAAGATATTCTTCTGAAAAATCTAAAGTA 9195
Qy      568  -----AlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuProSerProAspAla 585
Db      9196  GACCAAGTATTACACGATTTACAAAATGCTAAGAAT----- 9231
Qy      586  AspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyrAsnArgValLysGly 605
Db      9232  GACTTAAGAGCTGATAAAAGAGAAATTACAGACTGCA---TACAATAAATTA---TAC 9283
Qy      606  GluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrVal----- 622
Db      9284  AAAATCTTAATACCAATGCTGTA---AACCATCTAGTATTCAAACTAATAGTCTGCA 9340
Qy      623  ---GluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnLys 641
Db      9341  GACGAAATATCGAAACCAAT---ATAATACCGCT---AAAAATGAAGCACATAATGTTCTT 9396
Qy      642  PheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeu 661
Db      9397  -----GAAATACAAACCCCTACTGTAAAT-----GCAGTAGAAGAT--- 9432

```

```

Qy 662 PheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly 681
Db 9433 ---GCTTTACGTAAGATAAATGCAATTCACACGAGGTTTACA----- 9471
Qy 682 TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsnLys 701
Db 9472 -----AAAGCTATTATATCTTCAAGATTAAGAAGATAATAGCGAACTTGTAGACCA 9525
Qy 702 AsnPhelysAlaAspGlu-----GluProValGluGluThrProAlaGluPro 717
Db 9526 AAAGAAATATAGATCAAGCGATTAAATAGTCAACCATCACTAAATGGTATGACCTCAAGAA 9585
Qy 718 GluValProGlnValGluThrGluLysValGluAlaGln----- 730
Db 9586 TCTATTATAATATACACAAACAAACGTTAGAGAACACCAATTTAAACG-----GCAGAT 9756
Qy 731 -----LeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAla 748
Db 9646 ACTATTATATATATGGGATGCATCTATTGACAAATAACAGAAATAAATAATTCGAGTT 9705
Qy 749 ---AsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAsp 767
Db 9706 GAAGAGGCAACTAATGCATTTAACGAAGCAAAACCAACATTTAAACG-----GCAGAT 9756
Qy 768 AsnAsnSerIleMetAlaGluAlaGluLysLeuAlaLeuLysGlySerAsn--- 786
Db 9757 ACAACTTCTTAAAAAAGTACGTAACGAAATTTAAGTAGAGAGGCGCACAAACAAACAAA 9816
Qy 787 ---ProSerSerValSer 791
Db 9817 AAGCTAGAGTGTTAGT 9834

```

RESULT 22

US-08-956-171E-59  
 ; Sequence 59, Application US/08956171E  
 ; Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch  
 Gil H. Choi  
 Patrick S. Dillon  
 Craig A. Rosen  
 Steven C. Barash  
 Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 NUMBER OF SEQUENCES: 5256  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 31096 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 59:

US-08-956-171E-59

Alignment Scores:  
 Pred. No.: 0.00309 Length: 31096  
 Score: 152.50 Matches: 171  
 Percent Similarity: 32.80% Conservative: 137  
 Best Local Similarity: 18.21% Mismatches: 382  
 Query Match: 3.66% Indels: 249  
 Gaps: 39

US-09-765-271-56 (1-796) x US-08-956-171E-59 (1-31096)

```

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsn----- 15
Db 17213 TCACAGCTACAACTTAGCTGGTCTGAATCTGTTAAACAAATGCGAATAGTTAGAT 17272
Qy 16 -----AsnArgValSerTyrIleAspGlyLysGlnAlaThrGlnLysThrGlu 31
Db 17273 GGTGCTATGGCTAACTTACAAACGGCTATCAACGATAAGTCAGGAACATTAGCGAGCCAA 17332
Qy 32 AsnLeuThrProAspGluValSerLysArgGluGlyIleAsn-----Ala 46
Db 17333 AACTTCTTTGGATGCTGATGAGCAAAACGTAATGTCATCAATCAAGCTGTATCAGCAGCC 17392
Qy 47 GluGlnIleValIleLysIleThrAspGlnGlyTyrValThrSerHisGlyAspHisTyr 66
Db 17393 GAAACCATTTTAAATAAACAACACTGGA----- 17419
Qy 67 HisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSerGluGluLeuLeuMetLys 86
Db 17420 -----CCGAATACAGCGAAACAGCAGTCCGAAACAAAGATGCATTA 17458
Qy 87 AspProAsn-----TyrLysLeuLysAspGluAspIleValAsnGluValLysGly 103
Db 17459 AATGTTAATAATGCGAAACATGCATTAATGTTAGTACGAAAACTTAAACAAATCGGAACAA 17518
Qy 104 GlyTyrValIleLysValAspGlyLysTyrTyrVal-----TyrIleLysAspAlaAla 121
Db 17519 GCAGCGATTACAGCAATCAATCGCGCATCTGATTTAAATCAAAACAAAGATGCATTA 17578
Qy 122 HisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGln 141
Db 17579 AAAGCACAAGCTAATGCTGCTCAACGCGTATCTAATGCAAGATGTACAGCAACATCGG 17638
Qy 142 -----HisArgGluGlyThrProArgAsn-----AspGlyAlaValAla 155
Db 17639 ACTGAACGACGCGCAATGGGCATTTAAACATGCCATCGCAGATAAGCAGCAATACG 17698
Qy 156 LeuAlaArgSer-----GlnGlyArgTyrThrThrAsp 166
Db 17699 TTAGCAAGCAGTAAATATGTTAATGCCGATAGCAGCTTAAACAAATAATCTTACACAACTAA 17758
Qy 167 AspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVal 186
Db 17759 -----GTTACCAATGCTGAACATATTTAGCGGTACGCCAACCGTTGTAGACAA 17809
Qy 187 ProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeu--- 205
Db 17810 CCT-----TCAGAAAGTAACA 17824
Qy 206 AlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArg 225
Db 17825 GCTGCAGCTAATCAAGTAAACAGCGGAAACAGAAATTAATGTCGCGAAGATACGTT 17884
Qy 226 -----ArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerValSerAsn 243

```

Db 17885 GAAGCAAAACAAACGCC-----AATAGCTGCTATT 17914  
QY 244 ProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAlaSerGln 263  
Db 17915 GATGCAATTAACACAAATTAATTAACCTCAAAAGCTAAATTAAGAAGCAAGTGGGCA 17974  
QY 264 SerAsnAspIleAspSerLeuLeuLysGlnLeuTyLysLeuProLeuSerGlnArgHis 283  
Db 17975 GCCAATAGATTAGAAGCGTA-----CAAACT 18001  
QY 284 ValGluSerAspGlyLeuValPheAspProLalaGln-----IleThrSerArgThrAla 301  
Db 18002 GTTCAAAACAAATGGACAGCACTTGAACCAATGAAGGCTTAAGAGATAGTATTGCT 18061  
QY 302 ArgGlyValAlaValProHisGlyAspHisTyHisPheIleProTySerGlnMetSer 321  
Db 18062 AACGAACACACAGTCAAAACAGTCAAAACATATATACAGACGCAAGTCCGAATPAACCAATCA 18121  
QY 322 GluLeuGluGluArgIleAlaArgIleProLeuArgTyArgSerAsnHisTrpVal 341  
Db 18122 ACATATAATAGCGCTGCTCAAAATGCCAAGGTATCATTAATCAAACTAACATCCGACT 18181  
QY 342 ProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGlyPro 361  
Db 18182 ATGGATACTAGTGGGATTTACCCCAAGCTACAAACAGTGAATAATGCTTAAATAATGGTTA 18241  
QY 362 GlnProAlaProAsnLeuLysIleAspSerAsnSer----- 373  
Db 18242 AACGGTCTGAATACTTAAGAAATGCACAAACACACTGCTAAGCAAAACTTAAATPACATTA 18301  
QY 374 -----SerLeuValSerGlnLeuValArgLysValGlyGlu 385  
Db 18302 TCACACTTAACAAATAACCAAAATCTGCCATCTCATCACAAATGATCGGAGGTCAT 18361  
QY 386 GlyTyValPheGluGluLysGlyIleSerArgTyValPheAlaLysAspLeuProSer 405  
Db 18362 -----GTGAGTGAGGTAATCTGCTACTAATAAATGCAGCAACT 18397  
QY 406 Glu-----ThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSerValSer 422  
Db 18398 GATTGGAATACGCAATGGGTAACTTGGCAACAAAGCTATCCATGATCAAAACACAGTTAAA 18457  
QY 423 HisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyAspLys 442  
Db 18458 CAAAGTGTTAAATTTACTGATGCAGATAAGCTAAAGCTGAT----- 18499  
QY 443 AlaTyLys-----AsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArgAsn 461  
Db 18500 GCGTATACAAATCGGTAAAGCAGAGCTGAAGCAATTTCTGAATAAAACGCAAGGTGCAAT 18559  
QY 462 SerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsn----- 475  
Db 18560 ACCTCTAAACAGATGTTGAAGCGGTATTCAAAATGTTTCAAGTGCTAAAAATGCATTG 18619  
QY 476 -----AspGluSer-----ThrAsnLysGluLysLeuValAspAspLeuAlaPheLeu 492  
Db 18620 AATGGTGATCAAAACCTTACAAATGCGAAGATGCGAGCTTAAATGCAITTAATAACTTA 18679  
QY 493 AlaProIleThrHisProGluArgLeuGlyLysProAsn----- 505  
Db 18680 ACCTCAATTAATAATGCACAAAACGTCAGTAACTAACTAAATTTGATCAAGCAACAACT 18739  
QY 506 -----SerGlnIleGluTyThrGluAspGluValArgIleAlaGln----- 519  
Db 18740 GTAGCTGGTGTGAAGCTGTATCTAATACGAGTACACAAATGAATACAGCGATGGCTAAC 18799  
QY 520 -----LeuAlaAspLysTyThrThr-----SerAspGlyTyThrIlePheAsp 533  
Db 18800 TTGCAAAATGGTATTATGATATAAAACAAATACACTAGCAAGTGAACAACTAT----- 18850  
QY 534 GluHisAspIleLeuSerAspGluGlyAspAlaTyValThrProHisMetGlyHisSer 553  
|||||

Db 18851 ---CATGATGCTGATTTCAGATAAGAAAACCTGCTTATCTCAAGCGGTTACGAACGCGAA 18907  
QY 554 HisTrpIleGlyLysAspSerLeuSerLeuSerAspGlyGlyLysValAlaAlaGln----- 570  
Db 18908 AATATTTTAAATAAATAATAGTATCAATTTAGACAAAACCTGCGTTGAAACGCGTTG 18967  
QY 571 ---AlaTyThrLysGluLysGlyLysLeuProProSer-----ProAspAlaAsp 586  
Db 18968 TCACAAAGTGTGAATGCGAAAGGTGCCCTTAAATGGTAACTCAATAATTTAGACGAAGCTAAA 19027  
QY 587 ValLysAlaAsnProThrGlyAspSerAlaAlaIleTyAsnArgValLysGlyGlu 606  
Db 19028 TCATATGCAACACACTACT-----ATAACGCGACTT 19057  
QY 607 LysArgIleProLeuValArgLeuProTyMetValGluHisThrValGlyValLysAsn 626  
Db 19058 CAACATTTAAACACTGCTCAAAAGATAAATTTGAACAACAAAGTGCACACGACCAAAAT 19117  
QY 627 GlyAsnLeuIleIleProHisLysAspHisTyHisAsnIleLysPheAlaTrpPheAsp 646  
Db 19118 -----GTTGAGGTGTAGAT 19132  
QY 647 AspHisThrTyLysAlaProAsnGlyTyThrLeuGluAspLeuPheAlaThrIleLys 666  
Db 19133 -----ACTGTTAATCA---AGTGCCACACATTAATGGTGTATGGGTACGTTAAGA 19183  
QY 667 TyTrpValGlu-----HisProAspGluArgPro 676  
Db 19184 AATAGCATACAAGATAACACAGCTACGAAAAATGCGCAAAACTATCTTGTATGCTACAGAA 19243  
QY 677 HisSerAsnAspGlyTyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSer 696  
Db 19244 CGTAAACAAAACAACTATAACATGCTGTTGATAGTCTAATGGTGTATTAATGCAACA 19303  
QY 697 GluAspProAsnLysAsnPheLysAla-----Asp 706  
Db 19304 AGCAATCCAATATGATGCTGATGCTAATTAACCAATTCCTACACAAAGTGCATCAACG 19363  
QY 707 GluGluProValGluGluThrProAlaGluProGluValProGlnValGluThrGluLys 726  
Db 19364 AAAAAATGCAATGATGCTACACATAATTTAACGCAAGCGAAACAAACAGCAACAAATGCC 19423  
QY 727 ValGluAlaGlnLeuLys-----GluAlaGluValLeuLeuAlaLysValThr 742  
Db 19424 ATCGATGGTGCTACTAATTAATAAGCGCAAAAGATGCTTAAAGACCAAGTTACA 19483  
QY 743 AspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThr 762  
Db 19484 AGTGCGCAACGCTGTTGCAATGTAAACA---AGTATCCAAACAACTGCMAATGAACITTAAT 19540  
QY 763 -----LeuGln-----IleMetAspAsnAsnSer----- 770  
Db 19541 ACAGTATGGTCAATTAACACATGCTTATGATGATAAATGCAACAAACAACTCAA 19600  
QY 771 -----IleMetAlaGluAla 775  
Db 19601 AATATCGTACGCTGAACAAAGTAAGAAAACCTGCTTATGATCAAGCTGTAGTGTGCG 19660  
QY 776 GluLysLeuLeuAlaLeuLeuLysGlySerAsnProSerValSerLysGluLys 794  
Db 19661 AAAGCAATTTTAAATAAACAACAGGTTCAATTCAGATAAAGCAGGTTGACCGT 19717

## RESULT 23

US-09-417-197-108  
; Sequence 108, Application US/09417197  
; Patent No. 6518021  
; GENERAL INFORMATION:  
; APPLICANT: Ole THASTRUP, et al.  
; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An In  
; TITLE OF INVENTION: On A Cellular Response  
; FILE REFERENCE: 3759-0110P  
; CURRENT APPLICATION NUMBER: US/09/417,197  
; CURRENT FILING DATE: 1999-10-07

```
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 2616
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EGFP-Zap70 fusion
; NAME/KEY: CDS
; LOCATION: (1)..(2613)
US-09-417-197-108

Alignment Scores:
Pred. No.:      8,89e-05    Length:      2616
Score:          150.50     Matches:       148
Percent Similarity: 33.73%   Conservative:  107
Best Local Similarity: 19.58% Mismatches:    294
Query Match:        3.61% Indels:         209
DB:                4 Gaps:           31

US-09-765-271-56 (1-796) x US-09-417-197-108 (1-2616)

QY      62 HisGlyAspHisTyrHisTyrTyrsnGlyLysValProTyrAspAlalrleSerGlu 81
Dbb      232 CACATGAAGCAGCACGACTCTTCAAAGTCGCCCATGTCCCGAAGGCTAGCTCCAGGAGGC 291
QY      82 GluLeuLeuMetLysAspPro--asnTyrLysLeuLysAspGlu----- 95
Dbb      292 ACCATCTCTTTCAAGGACGGCAGCAAACAATAAACAGCCCCGCGCGAGTGTAAGTTCCAGGGC 351
QY      96 --AspllevalAnsluvalllys---glyGlyTyrValllleLysValaspGly---- 111
Dbb      352 GACACCCTGGTGAACCGCATCGAGCTGAAGGCGATCGACTTCAAGGAGACGGCAACATC 411
QY     112 -----LystyrtYrValtyrLeuLysAspAlala 121
Dbb      412 CTGGGGCCAAGAAGTGGAGTCAACAATAACAGCCCAACAAGCTATATCATGCGCCGACAAG 471
QY     122 HIsAlasPsnValArqThrylsGlulileAsnArglnLysGlnlHISer--- 140
Dbb      472 CAGAAGAAGCGGCATCAAGGTGAATTCAAGATCCGCCCAACATCATGAGACGGCAGCGTG 531
QY     141 ----GINHisARGgludgLyThrPrARgsnASpnSGlyAlaValAlaleuAla 157
Dbb      532 CAGCTCGCGGACCACTACCAGCAGAGAACCCCCTATCGGCGACGGCCCGCTGTCTGTG 588
QY     158 ArgSerGlnGlyArgTYrThrThrZspASPGLYTyrllEPheASNAlaserAspilelle 177
Dbb      589 -----CCCGACAAcCACTAcCTGAGCACCCAGTAGCAGCAGCCAGTCCGCCCTGAGC 627
QY     178 GluasPhTrGlyAspAlatyrilleValProHISglyAspHisTyrHistyrlieProLys 197
Dbb      628 AAAGACCCCAACGAG-----AAGCGC 648
QY     198 AsnGlnUeuSerAlaseRglueLuAlaLaIaGluAlaPheLseRGlyArgGlyAsn 217
Dbb      649 GATCATCATGGTCTCTGCTGAGTTCGTGACCGCCGCGCGGATCACTCTCGGCATGGACGAG 708
QY     218 LeuSerAnsSeraRgtThrTyrgargGlnAnsSerAspAsnThrSerArgthrAsntyp 237
Dbb      709 TTGTCAAGATCCGGACTCAGATCTCGAGTCTCAAGCTTCGAATTCGGCGATGCCAGAC-- 765
QY     238 ValProserValiserAsnpProgilyThrThrAsnThrSnThrSerAsnsenSerAsnthr 257
Dbb      766 ---CCCGCGCCCACTGCCC-----TTCTTCTACGCGCATCTCG 804
QY     258 AnSerGlnAlaSereSRansSpilEaspSerLeuleuLysGlnleutyrrLysLeu 277
Dbb      805 CGTGCAGGCGCGAGAGACCACTGAAGCTGGCGGGCATG----- 843
QY     278 ProLeuSerGlnARghisValGluSerASPGLYLuuValPheaSPproALaglNIlleThr 297
```

QY 604 sGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrValGluVa 624  
 Db 1877 ----- 1877  
 QY 624 lLysAsnGlyAsnLeuIleProHisGlyAspHisTyrHisAsnIleLysPheAlaTr 644  
 Db 1878 ----- 1878  
 QY 644 pPheAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaTh 664  
 Db 1913 AGGCGCATCAT-----GCACGAGCTGGCAACCCCTA----- 1946  
 QY 664 rIleLysTyrTyrValGluHis-----ProAspGluArgProHisSer-----As 679  
 Db 1947 -----CATCGTGGGCTCATTTGGCGTCTGCCAGCGCGGCCCTCATGCTGGTCATGG 1999  
 QY 679 nAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPr 699  
 Db 2000 AGATGGCTGGGGGGGGCGCTGCACAAAGTTCTCTGGTGGCAAGAGGAGGATCCCTG 2059  
 QY 699 oAsnLysAsnPhelysAla-----AspGluGluProValGluGl 712  
 Db 2060 TGACAATGTGGCGAGCTGCTGCACCAAGTGTCCATGGGATGAAGTACCTGGAGGAGA 2119  
 QY 712 uThr-----ProAlaGluProGluValPro 720  
 Db 2120 AGAACTTTGTCCACGTTGACCTGGCGGCCCGCAAGTCTCT 2159

RESULT 24  
 US-08-961-527-171  
 ; Sequence 171, Application US/08961527  
 ; Patent No. 6420135  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 171:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 14736 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 US-08-961-527-171

Alignment Scores: 0.00259 Length: 14736  
 Pred. No.:

Score: 148.00 Matches: 158  
 Percent Similarity: 33.54% Conservative: 109  
 Best Local Similarity: 19.85% Mismatches: 287  
 Query Match: 3.55% Indels: 242  
 DB: 4 Gaps: 38  
 US-09-765-271-56 (1-796) x US-08-961-527-171 (1-14736)  
 QY 107 lLysValAspGlyLysTyrTyrVal---TyrLeuLysAspAlaAlaHisAlaAspAsn 125  
 Db 9349 CTGAAATCGAAGGTATCAATATATTTGTTATATCAA---ACTAAGAAACAGATATAT 9405  
 QY 126 ValArgThrLysGluGluLeuAsnArgGlnLysGlnGluHisSerGlnHisArgGluGly 145  
 Db 9406 ACAGAGCTTTCAAGGACACAGTTGATGGGAATACTCTCTCAAGAGATAGTCAACCAAC 9465  
 QY 146 GlyThrProArgAsnAsp-----Gly 152  
 Db 9466 TCTACAAAACATCAGATGTAGTTCATTCACCTGATTAGATGGAAACCAAGGACAGGG 9525  
 QY 153 AlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsn 172  
 Db 9526 AAGGTAGTTTA-----CAAGGTCAAGCATCAGGGGATGATGGA----- 9564  
 QY 173 AlaSerAspIleIleGluAspThrGlyAspAlaTyrIleValProHisGlyAspHisTyr 192  
 Db 9564 ----- 9564  
 QY 193 HisTyrIleProLysAsnGluLeuSerAlaSerGluLeuAlaAlaGluAlaPheLeu 212  
 Db 9565 ---CTTTCAGAAAATCTCTATAGCAGCAGCAATCTATCTCTAATGATTCATTCGCA 9621  
 QY 213 SerGlyArgGlyAsnLeuSerAsnSerArgThrTyrArgGlnAsnSerAspAsnThr 232  
 Db 9622 AGT-----CAAGTTGACGAGATCCGATCACAA 9651  
 QY 233 SerArgThrAsnTrpValProSerValSerAsnProGlyThrThrAsnThrAsnThrSer 252  
 Db 9652 GGAGATCTGTAGTTCGACCAACAGTCCAGAACAGAGAAATCTGTCTCTCTACAACG 9711  
 QY 253 AsnAsnSerAsnThrAsnSerGlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLys 272  
 Db 9712 GTCAGAGTCGGAGAGAGTATTGCGCAGCAGCAAAATGAT-----CGA 9756  
 QY 273 GlnLeuTyrLysLeuProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAsp 292  
 Db 9757 CCAGAGTATAAACTTCCATTGGAAACCAAGGCGCAGCAAGAACCCGTCATGAGGTGAA 9816  
 QY 293 -----ProAlaGlnIleThrSerArgThrAlaArgGlyValAla 305  
 Db 9817 GCCGCGAGTCGTGAAGACTTACCAGTCTACACTAAGCCACTAGAAACCAAGGTACACAA 9876  
 QY 306 ValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMetSerGlnLeuGluGlu 325  
 Db 9877 GGACCCGGACATGAA-----GGTGAAGCTGCAGTTCGCGAGGAA 9915  
 QY 326 ArgIleAlaArgIleIleProLeuArg-----TyrArgSer 337  
 Db 9916 GAACACAGCTTACACAGAACCGTTAGCAACGAAAGGCGACGACAGCCAGTCTATGAGGC 9975  
 QY 338 AsnHisTrpValProAspSerArgProGluGlnProSerPro-----GlnProThr 354  
 Db 9976 AAAGCTACAGTCCGCGAAGAGACTCTTAGAGTACACGGAACCGTAGCGCAAAAGGCACA 10035  
 QY 355 ProGluProSerProGlyProGlnProAla-----ProAsnLeuLysIle 369  
 Db 10036 CAAAGAACCCGAACATCAGGCGGAGCGGAGAGAGAAAGAACTTCGCGCTTGAAGTTC 10095  
 QY 370 AspSerAsnSerSerLeuValSerGlnLeuVal----- 380  
 Db 10096 ACTACAGAAATAGAACGGAATCCAGATATTTCTTATACACAGAAATTCAGGAT 10155  
 QY 381 -----ArgLysValGly----- 384

```
Db 10156 CCAACACTTCTGAAAAATCGTGAAGATTGAACGACGAGGCAAGCGACACGCTACA 10215
QY |||||:
Db 385 -----GluGlyTyrValPheGluGluGlyGlySerArg 396
QY |||||:
Db 10216 ATTCATATGAGACTACATCGTAAATGTAATGTCGTAGAACTATAAGAGAGTGTCA 10275
QY |||||:
Db 397 TyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSer 416
QY |||||:
Db 10276 ACTGAAGTAGCT-----CCGCTCAACGAAGTCGTAAAGTAGGAACACTTGTGAAA 10326
QY |||||:
Db 417 LysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAsp 436
QY |||||:
Db 10327 GTTAAACCTCAGTAGAATAATCAACCTTAAACAAAGTTTGAAC----- 10371
QY |||||:
Db 437 GlnGluPheTyrAspLysAlaTyrAsnLeuLeu-----ThrGluAlaHisLysAlaLeu 454
QY |||||:
Db 10372 AAAAAATCTATAACTGTAACTTATACTTAAATAGACACTTACCTCAGCATATGTTCT--- 10428
QY |||||:
Db 455 Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArgLeu 474
QY |||||:
Db 10429 -----GCAAAACGCAAGTTTCCATGGA--GACAAGCTAGTTAAAGAGGTG 10473
QY |||||:
Db 475 AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro 494
QY |||||:
Db 10474 GATATAGAAATCTCTGCCAAGAGCAAGTAATATCAGGT---TTAGATTACTACACACG 10530
QY |||||:
Db 495 Ile-----ThrHis---ProGluArgLeuGlyLysProAsnSerGlnIleGluTyr 510
QY |||||:
Db 10531 TATACAGTTTAAACACACACTTAATATATTTGGGTGAAATAATAG-----GAAAT 10584
QY |||||:
Db 511 ThrGluAspGluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyr 530
QY |||||:
Db 10585 ACTGAACATCAACTCAAGATTCCCAATTA-----GAGTATAGAAATAAGAGATTAAA 10638
QY |||||:
Db 531 IlePheAsnGluHisAspIleLeuSerAspGluGlyAspAlaTyrValThrProHisMet 550
QY |||||:
Db 10639 GATATTGATTCAGTAGAATTATACGTAAGAAATGATCGTTAT----- 10683
QY |||||:
Db 551 GlyHisSerHisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGln 570
QY |||||:
Db 10684 -----CGTAGATATTTA-----AGTCTAAGTGAAGCGCGACTGATACGGCTAAA 10728
QY |||||:
Db 571 AlaTyrThrLysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsn 590
QY |||||:
Db 10729 TACTTTGTAAAGTGAAA----- 10746
QY |||||:
Db 591 ProThrGlyAspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIlePro 610
QY |||||:
Db 10747 -----TCAGATCGCTTCAAGAAATGTACTACCTGTAAATCT----- 10785
QY |||||:
Db 611 LeuValArgLeuProTyrMetValGluHisThr-----ValGluValLys 625
QY |||||:
Db 10786 -----ATTACAGAAATACGATGGAACGATATAAGTGCAGCGTAGCC 10827
QY |||||:
Db 626 AsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIleLysPheAlaTyrPhe 645
QY |||||:
Db 10828 GTTGATCAACTTCTCGAAGAGGTACAGCGTTACAAA----- 10866
QY |||||:
Db 646 AspAspHisThrTyr-----LysAlaProAsnGlyTyrThr--- 657
QY |||||:
Db 10867 GATGATTACATATTACTGTAGCTAAATCTAAAGCAGACGACCAAGGAGTTTACACATCC 10926
QY |||||:
Db 658 LeuGluAspLeuPheAlaThrIleLys-----TyrTyrValGluHisPro 672
QY |||||:
Db 10927 TTTAAACAGCTGTGAACAGCAAGCAATCTGTCTGGTGTCTATACATTGGCTTCA 10986
QY |||||:
Db 673 AspGluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys 692
QY |||||:
Db 10987 GATATG-----ACCGCATGAGGTGACCTTAGGCGAT 11019
QY |||||:
Db 693 LysAspHisSer-----GluAspProAsn 700
QY |||||:

Db 11020 AAGCAGACAAGTTATCTCACAGGTGCATTTTACAGGAGCTTGATCGGTTCTGATGGAACA 11079
QY |||||:
Db 701 LysAsnPheLysAlaAspGlu-----GluProValGluGluThrProAlaGluProGlu 718
QY |||||:
Db 11080 AAATCGTATGCCATTTATGATTGGAAGAACCATTAATTATTGATACATTAAATGCTGTCTACA 11139
QY |||||:
Db 719 ValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeu 738
QY |||||:
Db 11140 GTTAGAGATTGGATATTAACCTGTTTCTCTGATAGTAAAGAAAATGTCGACGCGTG 11199
QY |||||:
Db 739 AlalysValThrAspSerSerLeu-----LysAlaAsnAla 750
QY |||||:
Db 11200 CGCAAGGCGAGCAATAGCGCAATATTAATTAATGTCAGTAGAAGGAAAAATCTCAGGT 11259
QY |||||:
Db 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnSer 770
QY |||||:
Db 11260 CGCAAAATCTGTTGCGGATTTAGTCGAGCGCAACAATAACATACAGTAGATAAAACAGCTCG 11319
QY |||||:
Db 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySerAsn 786
QY |||||:
Db 11320 TTTACAGGG-----AACTTATCGCAATCACCAGGACAGTAAT 11358

RESULT 25
US-09-817-514A-5
; Sequence 5, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2745
; TYPE: DNA
; ORGANISM: Photorhabdus luminescens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2745)
US-09-817-514A-5

Alignment Scores:
Pred. No.: 0.000191 Length: 2745
Score: 147.50 Matches: 178
Percent Similarity: 32.99% Conservative: 112
Best Local Similarity: 20.25% Mismatches: 380
Query Match: 3.54% Indels: 210
DB: 4 Gaps: 35

US-09-765-271-56 (1-796) x US-09-817-514A-5 (1-2745)
QY 13 LysGluAsnAsnArgValSerTyrIleAsp-GlyLysGlnAlaThrGlnLysThrGluAs 32
Db 408 AGAGAAAACGACGGAACGCTTTATCTGGCGCGCAATACGCCGCAAGAAAAGATACAA 467
QY 32 nLeuThrProAspGluValSerLysArgGlu----- 42
Db 468 CCTCGCGGTCACTGTGTCCGCCATTACGATACCGCGGAGCTTACTCACTCAATAGCCT 527
QY 43 -----GlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVa 59
Db 528 TTCTCTGGTGGCGTGTGCTATCAAACTCAACACTCTCAACACTCTCCTACCATCAACAGGATGC 587
QY 59 lThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleI 79
Db 588 CGACTGCAGGTTGAGACACGAGCCCTCTGGCAACAAAACACTGATGATGTCTATAT 647
```

QY	79	eSerGluGlu-----LeuLeuMetLysAspPro-----As	89
Db	648	CAACCAAGTAACACTGATGCCACCGGGCTTTACTGACCCAGACCGATGCCAAAGGCAA	707
QY	89	nTyrLysLeuLysAspGluAspIleValAsnGluValLysGlyGlyTyrValIleLysVa	109
Db	708	CATTCAAGCGCTGGCTATGATGTGGCGGCGAGCTAAAGGGAGTTGGTTAACTCAAC	767
QY	109	lAspGlyLysTyrTyrValTyrLeuLysAspAlaIleHis-----AlaAspAsnValar	127
Db	768	AGTCAGCGCGAACAGGTGATTAATAATCGTAACCTACTCCGCGCGGCGCAAAAAATT	827
QY	127	gThrLysGluGluIleAsnArgGlnLysGlnGluHisSerGlnHisArgGluGlyGlyTh	147
Db	828	ACGTGAAGACGACGGTAAACGGGATTGTCAGTGAATACAGCTACGAAACCGGAA-----AC	881
QY	147	rProArgAsnAspGlyAlaValAlaLeuAlaArgSerGlnGlyArgTyrThrThrAspAs	167
Db	882	COAACGGCTTATCGGCATTACCACTCGCGCTCCATCAGACGCCAAGGTTGCGAAGCCT	941
QY	167	pGlyTyrIlePheAsn---AlaSerAspIleIleGluAspThrGlyAspAlaTyrIleVa	186
Db	942	ACGCTATCAATATGACCAGCGTAGCAATGTCTAATAATCCGTATATGATCGGAAGCCAC	1001
QY	186	lProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeuSerAlaSerGluLeu--	205
Db	1002	TCGCTTTTGGCGCAATCAGAAAGTAGCCCGGAGAAATAGCTATACCTACGATTCCCTGTA	1061
QY	206	-----AlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsnSerAr	222
Db	1062	TCAGCTTATCAGCGCCACCGGCGCGAAATGCGCAATATCGTCAGCAAAACAACCAACT	1121
QY	222	gThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSerVal--	241
Db	1122	TCCCTCCCTCGGTACTCTTGACAACAATCACTACATACTATCTACTCGCACTACAG	1181
QY	242	-----SerAsnProGlyThrThrAsn--	248
Db	1182	CTATGATCACAGTGTAATCTGACGCAAAATTCGGCACAGCTCGCCAGCTACCCAGAACAA	1241
QY	249	-----ThrAsnThrSerAsnAsnSerAsnThrAsnSerGlnAl	261
Db	1242	CTACACCGTGGCTATCACCCCTCTCAAAACCGCAGCAATCGGGGTCTCTCAGTACGCTAAC	1301
QY	261	aSerGlnSerAsnAspIleAspSerLeuLeu-----LysGlnLeuTyrLysLe	277
Db	1302	CACGATCCAAATCAAGTGGATAGCTGTTGTGACCGGTGGTGTCACCAAAACCACTTTATT	1361
QY	277	uProLeuSerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleTh	297
Db	1362	ACCC-----GGACAGACACTTATCTGGACACCCACGAGAGAGTT	1400
QY	297	rSerArgThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTy	317
Db	1401	AAAGCAGGTTAATAATGGC-----CGGGAAATAGTGGTAC-----	1437
QY	317	rSerGlnMetSerGlnLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSe	337
Db	1438	-----CGCTACGACACAG	1448
QY	337	rAsnHisTrpValProAspSerArgProGluGlnProSerProGlnProThrProGluPr	357
Db	1449	CAACGGCATGAGACAACCTGAAGATGAGTGAACAGGCCAACCCAGAAATACTACGAGCAACA	1508
QY	357	oSerProGlyProGlnProAlaProLeuLeuLys---IleAspSerAsnSerSerLeuVa	376
Db	1509	ACGGTAAATCTATTTCGGGAGCTGGAGCTACGCACAACCCAGAGCAACGCCACAACAAC	1568
QY	376	lSerGlnLeu---ValArgLysValGlyGlyGlyTyrValPheGluGluLysGlyIleSe	395
Db	1569	GGAAAGATTACAGTTATCACTCGGTGAAGCC-----GGTGGCGCAGGT	1616

QY	395	rArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLe	415
Db	1617	ACGGGTGTTGCACCTGGGAGAGCGGTAAAGCAGAGATGTCACCAATAATCAACTACGTTA	1676
QY	415	uSerLysGlnGluSerVal---SerHisThrLeuThrAlaLysLysGluAenValAlaPr	434
Db	1677	CAGCTACGATATCTGATCGGCTCCAGCCAGCTTGAATCGACCAACCAAGGACACAAATAT	1736
QY	434	oArgGlnGlnPheTyrAsp-----LysAlaTyrAsnLeuLeuTh	448
Db	1737	CAGCCGAGGAGATATATCATTTGGCGGCACAGCGCTGGGCACGACCAACAGCCAAAC	1796
QY	448	rGluAlaHisLysAlaLeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAs	468
Db	1797	AGAACCAGCTATAAAACGATTCGCTATTCGGCAAAGACGAGATGCCACCGGTTGTA	1856
QY	468	pLysLeuLeuGluArgLeuAenAspGlu-----SerThrAsnLysG1	482
Db	1857	TTATTACGGTTATCGTTATTACCAACCGTGGCGGCAGATGTTAAGCCGCGACCGCGC	1916
QY	482	uLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArgLeuG1	502
Db	1917	AGGAACCATTCATGGCTGAATCTATACCGAATGGTAAGAAATAATCCTGTGAGTTA--	1974
QY	502	yLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeuAlaAs	522
Db	1975	-----CAAGATGAATGGATTAGCGCCAGAAAGG	2006
QY	522	pLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAspGluG1	542
Db	2007	GAATATACCAAGAGGTAATTTC---TTTGATGAATTAATAATCAAATTGGCAGCCAA	2063
QY	542	yAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSerLeuSe	562
Db	2064	AAGTTCACATGTTGTC-----AANTGCAACGAGAAAGAGAC-----	2100
QY	562	rAspLysGlnLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLe---LeuProPr	581
Db	2101	-----AGTTATACAAAAATAATCATTTGAAAGTGTTGTCG	2135
QY	581	oSerProAspAlaAspValLysAlaAsnProThrGly-----	593
Db	2136	TGTCGTGATTCGAT-----CGTCGGGTATTGCTAAGCCACGAGAGTT	2183
QY	594	-----AspSerAlaAlaIleTyrAsnArg-Val-----	603
Db	2184	ACTAAAAGGTATAGAAAAAGTCAATATATATACCGCATTTGAAGAAAAACAGTCCCT	2243
QY	604	-----LysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHisThrV	622
Db	2244	TTCAGAAAAATCAAAACGAAATCTTTCTTTAGATCTGAAATATCGGTTATATGCAAG	2303
QY	622	alGluValLysAsnGlyAenLeuIleIleProHisLysAspHisTyrHisAsnIleLysP	642
Db	2304	AA-----CCATACAAGATACGATATCAGATATGTC-----	2334
QY	642	heLaTrpPheAspAspHisThrTyrLysAla-----ProAsnGlyTyrThrLeuGluA	660
Db	2335	-----GAAGACATAAATAGAAATATCACCCTGATTTTATTCAGAAACCG	2383
QY	660	spLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnA	680
Db	2384	ATTCTCTTCGTTAATG-----GATAAAGTGAATAAATG	2419
QY	680	spGlyTrpGlyAsnAlaSerGluHis---ValLeuGlyLysLysAspHisSerGluAspP	699
Db	2420	ATTATTCGGGTGAAGAAAAATTTATGGCGCAATCGAGGTTAAGGTTTATCATGATTTAA	2479
QY	699	roAsnLysAsnPheLysAlaAspGluGluProValGluThrProAlaGluProGluV	719
Db	2480	AAAATAAACAA-----TCAGAAATTACGTCACTATGCATGTGCCATCCC-----	2526
QY	719	alProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuA	739



Fri Oct 1 15:35:51 2004

```

Db 2527 -----TATACCGCAATTGAGTAATGAACAAAGAGCGCTGT 2560
Qy 739 laLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgA 759
Db 2561 TGCAGAAACACAGAACCCGCTATTGCAATAGATAGAGAATATAATTTCAAAGGTGTTGGCA 2620
Qy 759 snAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuL 779
Db 2621 AATTCTCGACAATGAAGCAATTAATAATCA----- 2652
Qy 779 euAlaLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db 2653 -----TTGAAGGACATAAAATTAAATAGGATATCAACAGAGGCTATTAAAT 2697

```

Search completed: October 1, 2004, 11:34:14  
Job time : 363 secs

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      October 1, 2004, 10:19:33 ; Search time 928 Seconds
           (without alignments)
           4348.465 Million cell updates/sec

Title:       US-09-765-271-56
Perfect score: 4165
Sequence:    1 SYEGLYQARTVKENRVS.....KLLALLKGNSPSVSKIN 796

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched:   3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters:      6681306

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 90 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09765271/runat_30092004_113756_12374/app_query.fasta_1.967
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=90 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO.MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LOGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result      *
No.         Query      Match Length DB ID      Description

```

1	4163	100.0	2389	9	US-09-765-272-55	Sequence 55, Appl
2	4163	100.0	2451	13	US-10-412-862-9	Sequence 9, Appl
3	4163	100.0	2451	16	US-10-412-850-9	Sequence 9, Appl
4	4163	100.0	2451	16	US-10-387-783-9	Sequence 9, Appl
5	4163	100.0	8195	13	US-10-158-844-94	Sequence 94, Appl
6	3218	77.3	2523	10	US-09-884-465A-3	Sequence 3, Appl
7	3218	77.3	2647	10	US-09-884-465A-4	Sequence 4, Appl
8	2795	67.1	2481	10	US-09-769-787-206	Sequence 206, App
9	2786.5	66.9	2531	13	US-10-412-862-11	Sequence 11, Appl
10	2786.5	66.9	2531	16	US-10-412-850-11	Sequence 11, Appl
11	2786.5	66.9	2531	13	US-10-387-783-11	Sequence 11, Appl
12	2772	66.6	2531	16	US-10-412-862-5	Sequence 5, Appl
13	2772	66.6	2531	16	US-10-412-850-5	Sequence 5, Appl
14	2772	66.6	2531	16	US-10-387-783-5	Sequence 5, Appl
15	2750.5	66.0	2639	10	US-09-884-465A-5	Sequence 5, Appl
16	2649.5	63.6	2290	9	US-09-765-272-65	Sequence 65, Appl
17	1774	42.6	2359	13	US-10-158-844-243	Sequence 243, App
18	1247	29.9	3120	10	US-09-884-465A-1	Sequence 1, Appl
19	1247	29.9	5048	10	US-09-884-465A-2	Sequence 2, Appl
20	1222.5	29.4	6867	13	US-10-158-844-192	Sequence 192, App
21	1203.5	28.9	1455	10	US-09-769-787-246	Sequence 246, App
22	1203.5	28.9	1455	10	US-09-769-787-246	Sequence 23, Appl
23	1203.5	28.9	1455	13	US-10-412-862-7	Sequence 23, Appl
24	1203.5	28.9	1455	16	US-10-412-850-7	Sequence 7, Appl
25	1203.5	28.9	1455	16	US-10-387-783-7	Sequence 7, Appl
26	1187.5	28.5	1342	9	US-09-765-272-181	Sequence 181, App
27	988	23.7	973	13	US-10-158-844-355	Sequence 355, App
28	966.5	23.2	2528	10	US-09-884-465A-9	Sequence 9, Appl
29	946	22.7	5215	10	US-09-252-088-13	Sequence 13, Appl
30	946	22.7	5215	16	US-10-340-792-13	Sequence 13, Appl
31	929	22.3	2469	10	US-09-769-736-17	Sequence 17, Appl
32	841	20.2	1146	10	US-09-769-736-23	Sequence 23, Appl
33	770.5	18.5	1684	13	US-10-158-844-258	Sequence 258, App
34	650	15.6	2721	16	US-10-324-143-6	Sequence 6, Appl
35	644	15.5	816	16	US-10-324-143-4	Sequence 4, Appl
36	644	15.5	816	16	US-10-324-143-5	Sequence 5, Appl
37	643	15.4	819	10	US-09-884-465A-257	Sequence 257, App
38	643	15.4	819	16	US-10-324-143-10	Sequence 10, Appl
39	512.5	12.3	1455	10	US-09-769-736-71	Sequence 71, Appl
40	304.5	7.3	841	9	US-09-452-599-34	Sequence 34, Appl
41	304.5	7.3	841	15	US-10-121-120-34	Sequence 34, Appl
42	172	4.1	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
43	170	4.1	7434	9	US-09-815-242-4761	Sequence 4761, Ap
44	170	4.1	7437	9	US-09-815-242-8869	Sequence 8869, Ap
45	170	4.1	8155	8	US-08-781-986A-63	Sequence 63, Appl
46	170	4.1	8155	13	US-10-329-624-63	Sequence 63, Appl
47	169	4.1	2730	15	US-10-172-502-5	Sequence 53, Appl
48	169	4.1	12173	8	US-08-781-986A-310	Sequence 310, App
49	169	4.1	12173	13	US-10-329-624-310	Sequence 3, Appl
50	168.5	4.0	3132	9	US-09-946-805-3	Sequence 7677, Ap
51	164	3.9	2676	13	US-10-282-122A-7677	Sequence 35051, A
52	163	3.9	11076	13	US-10-282-122A-35051	Sequence 7, Appl
53	159.5	3.8	14172	16	US-10-359-012-7	Sequence 4438, Ap
54	159	3.8	2613	9	US-09-815-242-4438	Sequence 8539, Ap
55	159	3.8	2613	9	US-09-815-242-8539	Sequence 34396, A
56	159	3.8	19922	13	US-10-282-122A-34396	Sequence 94, Appl
57	158	3.8	4359	10	US-09-873-367C-94	Sequence 103, App
58	158	3.8	4359	15	US-10-153-668-103	Sequence 3, Appl
59	157	3.8	5419	17	US-10-479-546-3	Sequence 103, App
60	157	3.8	7088	17	US-10-451-467A-383	Sequence 383, App
61	156	3.7	1766	10	US-09-934-455-163	Sequence 163, App
62	156	3.7	1766	13	US-10-412-699B-697	Sequence 697, App
63	156	3.7	1766	13	US-10-225-066A-599	Sequence 599, App
64	156	3.7	1766	16	US-10-302-267-35	Sequence 35, Appl
65	156	3.7	1766	16	US-10-374-780A-2519	Sequence 2519, Ap
66	155.5	3.7	3369	13	US-10-282-122A-34397	Sequence 34397, A
67	154	3.7	5085	15	US-10-198-846-9854	Sequence 9854, Ap
68	153.5	3.7	5967	15	US-10-171-581-158	Sequence 158, App
69	152.5	3.7	7200	13	US-10-282-122A-33992	Sequence 33992, A
70	152.5	3.7	31096	8	US-08-781-986A-59	Sequence 59, Appl
71	152.5	3.7	31096	13	US-10-329-624-59	Sequence 59, Appl
72	151.5	3.6	6258	9	US-09-815-242-4584	Sequence 4584, Ap

Sequence 8512, Ap  
Sequence 1, Appli  
Sequence 108, App  
Sequence 1, Appli  
Sequence 7653, Ap  
Sequence 7627, Ap  
Sequence 171, App  
Sequence 5, Appli  
Sequence 26, Appl  
Sequence 5, Appli  
Sequence 32569, A  
Sequence 45972, A  
Sequence 4580, Ap  
Sequence 8291, Ap  
Sequence 43, Appl  
Sequence 33842, A  
Sequence 33993, A  
Sequence 38279, A

# ALIGNMENTS

RESULT 1  
US-09-765-272-55  
; Sequence 55, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-765-272-55  
Alignment Scores:  
Pred. No.: 0 Length: 2389  
Score: 4163.00 Matches: 796  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.95% Indels: 0  
DB: 9 Gaps: 0

US-09-765-271-56 (1-796) x US-09-765-272-55 (1-2389)  
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
Db 2 TCTTACGAGTGGGACTGCTATCAAGCTAGAAGCGTTAAGGAAAATAATCGTGTTCCTAT 61  
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
Db 62 ATAGATGAAAACCAAGCGACGCAAAACGAGAAATTTGACTCTCTGATGAGGTTAGCAAG 121  
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
Db 122 CGTGAAGAAATCAATGCTGAGCAATCTCATCAAGATTAACAGACCAAGGCTATGCTACT 181  
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80  
Db 182 TCACATGCGGACCACTATCATTTTACATGTTAGGTTTCTTATGACGCTATCATCAGT 241  
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100  
Db 242 GAAGAATTAATCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGGATTTGTTAATGAG 301  
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrValLysAspAla 120  
Db 302 GTCAAGGGTGGATATGTTATCAAGGTAGTAGGAAAATACTATGTTTACCTTAAGGATGCT 361  
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
Db 362 GCCACGCGGATAACGTCGGTACAAAAGAGGAAATCAATCGACAAAAACAAGAGCATAGT 421  
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaIleAlaArgSerGln 160  
Db 422 CAACATCGTGAAGTGGAACTCCAAGAAACGATGCTGCTGTTGCCCTGGCAGCTTCGCAA 481  
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
Db 482 GGACGCTATPACTACAGATGATGTTATATCTTTATGCTTCTGATATCATAGAGGATACT 541  
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
Db 542 GGTGATGCTTATATGCTTCTCATGAGATCATTAACATTACATTCTAAGAATGAGTTA 601  
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyValAsnLeuSerAsn 220  
Db 602 TCAGCTAGCGAGTGGCTGCTCAGAAAGCTTCTCTATCTGGTCGAGAAATCTGTCAAT 661  
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
Db 662 TCAAGAACTTATCGCCGACAAATACGATACACTTCAAGAACAACTGGGTACCTCT 721  
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
Db 722 GTAAGCAATCCAGGAACACTCAAACTATAACAAAGCAACAAACAGCAACACTAACAGTCA 781  
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGluLeuTyrLysLeuProLeuSer 280  
Db 782 GCAAGTCAAAAGTAAATGACATTTGATGCTCTTGAACACAGCTCTACAAACTGCTTTGAGT 841  
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
Db 842 CAACGACATGTAGAAATCTGATGCTTGTCTTGTATCCAGCACAAATCACAAGTCGAACA 901  
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
Db 902 GCTAGAGGTGTTGAGTGGCCACACGAGATCATTAACACTTCTATCCCTTACTCTCAATG 961  
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
Db 962 TCTGAATGGAAGAACGAATCGCTCGTATTATTTCCCTTGTGTTATCGTTCAACCATGG 1021  
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
Db 1022 GTACCAGATTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1081

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1082 CGCAACCTGCACCAAAATCTTAAATAGACTCAATCTCTTTGGTGGTGGTGA 1141  
 QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1142 CGAAAGTTGGGAAGGATATGTTTCGAAGAAAGGGCATCTCTGTTATGCTTTGGC 1201  
 QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 1202 AAAGATTTACCAATCTGAAACTGTTAAATCTTTGAAAGCAAGTTATCAAAAACAAGAGAT 1261  
 QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1262 GTTTCACACACTTTAACTGCTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTAT 1321  
 QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1322 GATAAAGCATATAATCTGTTAACTGAGGCTCATAAGGCTTGTGTTGNAATAAGGTCGT 1381  
 QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuArgLeuAsnAspGluSerThrAsn 480  
 Db 1382 AATCTGATTTCCAAAGCTTTAGACAAATATTAGAACGCTTGAATGATGAATCGACTAAT 1441  
 QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1442 AAAGAAATAATGGTAGATGATTTATTGGCAATCTCTAGCACCAATYACCCATCCAGAGCA 1501  
 QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1502 CTTGGCAAAACCAAAATCTCAAAATGAGTATATCTGAAGACGAATTCGTATTGCTCAATTA 1561  
 QY 521 AlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1562 GCTGATAGTATACACGTCAGATGGTTACATTTTGGATGAACATGATATAATCAGTGAT 1621  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 1622 GAAGGAGATGCATATGTAAGCCCTCATATGGCCATAGTCACTGGATGGAAAGATAGC 1681  
 QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1682 CTTTCTCATAGGAAAGAAAGTTGACGCTCAAGCCTATATCTAAAGAAAGGATATCCTACCT 1741  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1742 CCATCTCCAGACGAGATGTTAAGCAAAATCCAACTGGAGATAGTCAGCAGCTATTTTAC 1801  
 QY 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1802 AATCGTGTGAAGGGGAAACCAATTCCTGCTGCTGCTTCCATATATGGTTGACAT 1861  
 QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1862 ACAGTTGAGGTTTAAACCGGTAAATTTGATTAATTTCTCATAGGATCATTAACCAATAT 1921  
 QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1922 AAATTTGCTGTTTGTGATGATCACATACAAAGCTCCAAATGGCTATATCCTTGGAGAT 1981  
 QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 1982 TTGTTTGGCAGGATTAAGTACTAGTAGAACACCCCTGACCAAGCTCCACCATTTATGAT 2041  
 QY 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2042 GGATGGGGCAATCCGATGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2101  
 QY 701 LysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2102 AAGAACTTCAAGCGGATGAAGCAGTAGAGGAAACACCTCTGAGCAGGAAGTCCCT 2161

QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740  
 Db 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCAACCTCAAGAGACGAGAAAGTTTGTCTTGGAAA 2221  
 QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2222 GTAACGGGATTCAGTCTGAAAGCCAATGCAACAGAAACTCTAGCTGGTTTACGAATAAT 2281  
 QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuAla 780  
 Db 2282 TTGACTCTTCAATATATGATAACAATAGTATCATATGGCAGAGCAGAAAAATTTACTTGGC 2341  
 QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 2342 TTGTTTAAAGGAAGTAATCTCTCATCTCTAAGTAGGAAAAATAAAC 2389  
 RESULT 2  
 US-10-412-862-9  
 ; Sequence 9, Application US/10412862  
 ; Publication No. US20040052781A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-685  
 ; CURRENT APPLICATION NUMBER: US/10/412,862  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 9  
 ; LENGTH: 2451  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(2451)  
 ; OTHER INFORMATION: n = a, c, t or g  
 US-10-412-862-9  
 Alignment Scores:  
 Pred. No.: 0 Length: 2451  
 Score: 4163.00 Matches: 796  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.95% Indels: 0  
 DB: 13 Gaps: 0  
 US-09-765-271-56 (1-796) x US-10-412-862-9 (1-2451)  
 QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 61 TCTTACGAGTTGGAGCTGATCAAGCTTAGCGAAATAATCGTGTTCCTAT 120  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 121 ATAGATGGAAACACACGACGCAAAACCGGAAATTTGACTCTCGATGAGGTAGCAAG 180  
 QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 181 CGTGAAGGAATCAATCTCGACCAATCGTCATCAAGATAACAGCCAGCTATGTCAC 240  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80  
 Db 241 TCACATGGCGACCATCATATTATTAACAATGGTAAGGTTCTTATCATCGCTATCATCAGT 300  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100

Db 301 GAAGAAATTCATCGAAGATCCAAACTATTAAGCTAAAGATGAGGATATTGTTAATGAG 360  
 Qy 101 VallysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 361 GTCRAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGCT 420  
 Qy 121 AlahisAlaAspAsnValArgThrLysGluGluLysLeuArgGlnLysGlnLysHisSer 140  
 Db 421 CCCCACGCGGATAAGCTCGTACAAAGAGGAAATCAATCGACAAAAACAAGAGATAGT 480  
 Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 481 CAACATCGTAGGAGGTGGAACTCCAAAGAAACGATGCTGTGTGCTTGGCAGTTCCGAA 540  
 Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 541 GGACCTATACATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATACT 600  
 Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 601 GGTGATGCTTATATCGTTCCTCATGAGATCAATTACCATTAATCTCCTAAGAAATGAGTTA 660  
 Qy 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 661 TCAGCTAGCGAGTTGGCTGCTGCAGAGCTTCTCTATCTGTCGAGGAAATCTGTCAAAAT 720  
 Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240  
 Db 721 TCAAGAACCTATACGCCGCAAAATAGCGATAACACTTCAAGAACAACTGGGTACTCTCT 780  
 Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 781 GTAAGCAATCCAGGAACTCAAAATATAACACAGCAACACAGCAACACTAACAGTCAA 840  
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 841 GCAGTCAAGTAATGACATTGATGCTCTTTGAAACAGCTCTACAACTGCTTTGAGT 900  
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 901 CAACGACATGTAAATCTGATGCTGCTTTGATCCAGCAAAATCAAAAGTCGAAACA 960  
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 961 GCTAGAGGTGTCAGTGCACACGAGATCATACACTTATCCCTTACTCTCAATG 1020  
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 1021 TCTGAATTCGAAGACGAATCGCTCGTATTATTCCCTTCGTTATCGTTCAAAACCATGG 1080  
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 1081 GTACCAAGATTCAAGGCGCAGAACCAACCAAGTCCAAACCGACTCCGGAACCTAGTCAGGC 1140  
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1141 CCGCAACCTGCACCAAACTTTAAATAGACTCAAAATCTCTTTGGTTAGTCACTGTA 1200  
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1201 CGAAAGTTGGGAGGAGATATGATTTCGAAGAAAAGGGCATCTCTCGTTATGCTTTGCG 1260  
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLysLeuSerLysGlnGluSer 420  
 Db 1261 AAAGATTATACCATCTGAACCTGTTAAATAATCTTTGAAAGCAAGTTATCAAAACCAAGAGT 1320  
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1321 GTTTTACACACTTTAATGCTTAAAGAAAAGAAATGTTGCTCTCTCGTACCAAGAAATTTAT 1380  
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1381 GATAAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTGTTTGNAAAATAAGGGTCT 1440

Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1441 AATTCTGATTTCCAAAGCCTTAGACAAATATTAGAACGCTTGAATGATGAATCAGCTAAT 1500  
 Qy 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1501 AAAGAAAATTTGGTAGATGATTATTGGCATTCCTAGCACCAATTTACCCAGAGCGA 1560  
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1561 CTTGGCAAAACCAAAATCTCAAAATTGAGTATACTGAAGACGAAGTTCTGTTTGTCTCAATTA 1620  
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1621 GCTGATAGTATATCAACGTCAGATGCTTACATTTTGTATGAACATGATATAATCAGTCAT 1680  
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 1681 GAAGGAGATGCATATGTAACGCTCATATGGGCCCATAGTCACTGGATTGGAAAAGATAGC 1740  
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1741 CTTTCTGATAGGAAAAGTTGCGACTCAAGCTATATCTAAGAAAAAGGTATCTCTACT 1800  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 1801 CCATCTCCAGAGCGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTTAC 1860  
 Qy 601 AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis 620  
 Db 1861 AATCGTGTGAAAGGGGAAAACGAAATCCACTCGTTCGACTTCCATATATGTTGAGCAT 1920  
 Qy 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640  
 Db 1921 ACAGTTGAGTTTAAAAACGTAATTTGATTTCTCTCATAGATCATTTACCATATATTT 1980  
 Qy 641 LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660  
 Db 1981 AATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTATACCTTGAAGAT 2040  
 Qy 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680  
 Db 2041 TTGTTTGGCCACCATTAAGTACTACCTAGAACACCCCTGACGAACCTCCACATTTCAATGAT 2100  
 Qy 681 GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 2101 GGATGGGCAATGCCAGTGCAGCATGTGTTAGGCAAGAACACACAGTGAAGATCCAAAT 2160  
 Qy 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720  
 Db 2161 AAGAACTTCAAAGCGGATGAAGAGCCAGTAGAGAAACACCTCTGAGCCAGAAAGTCCCT 2220  
 Qy 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuAlaLys 740  
 Db 2221 CAAAGTAGAGACTGAAAAGTAGAAGCCCACTCAAGAGACGAGAGTTTTGCTTGCAGAA 2280  
 Qy 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760  
 Db 2281 GTAAAGGATTTCTAGTCTGAAGCCATGCAACAGAACTCTAGCTGGTTTACCAATAAT 2340  
 Qy 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780  
 Db 2341 TTGACTCTTCAAATTTATGGATAACAATAGTATCATGCGAGAGACAGAAAATTTACTTGGC 2400  
 Qy 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796  
 Db 2401 TTGTTAAAGGAAGTAACTCTTCTATCTGTAAAGTAAGGAAAATAAATTAAC 2448

RESULT 3  
 US-10-412-850-9  
 ; Sequence 9, Application US/10412850  
 ; Publication No. US20040001836A1

```

; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) - (2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-412-850-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 16 Gaps: 0

US-09-765-271-56 (1-796) x US-10-412-850-9 (1-2451)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
DB 61 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTTTCCTAT 120
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 121 ATAGATGGAAACAAAGGAGCGCAAAACCGAGAAATTTGACTCTCGTATGAGGTTAGCAAG 180
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
DB 181 CGTGAAGGAATCAATGCTGAGCAATCGTCAATCAAGATACAGACCAGGCTATGTCACT 240
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80
DB 241 TCACATGGCGACCACTATCATATTACAAATGGTAAGGTTCTTATGACGCTATCATCAGT 300
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
DB 301 GAAGATTAATCATGAAGATCCAAACTATAAGCTAAAGATGAGGATATGTTAATGAG 360
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
DB 361 GTCAAGGCTGGATATGTTATCAAGTAGATGGAAATACTATGTTTACCTTAAGGATGCT 420
QY 121 AlaHisAlaAspAsnValArgThrLysGluLysIleAsnArgGlnLysGlnHisSer 140
DB 421 GCCCACGGATACGTCCTGACAAAGAGGAAATCAATCAACAAACAAAGACATAGT 480
QY 141 GlnHisArgGluGlyThrProAspAsnGlyAlaValAlaLeuAlaArgSerGln 160
DB 481 CAACATCGTGAAGTGAATCCAAAGAAACGATGGTGTCTGCTTGGCACGTTGCGAA 540
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 541 GGACGCTATACACAGATGATGTTATCTTAAATGCTTCTGATATCATAGAGGATACT 600
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

```

```

DB 601 GGTGATGCTTATATCGTTCCTCATCGAGATCATTTACCATTACATTCCTTAAGAAATGAGTTA 660
QY 201 SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 661 TCAGCTAGCGAGTTGGCTGCTGCAGAACCTTCTTATCTGCTCGAGGAAATCTGTCAAT 720
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
DB 721 TCAGAAGCTATCGCCGCAAAATAGCATACACTTCAAGAACCAACTGGGTGCTTCT 780
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
DB 781 GTAAGCAATCCAGGAACTTCAAAATACTAACACAGCAACACAGCAACACTAACAGTCAA 840
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
DB 841 GCAAGTCAAAGTAATGATGATGATGCTCTTGAACAGCTCTACAAACTGCCCTTGAGT 900
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
DB 901 CAACGACATGTAGAAATCTGATGGCTTGTCTTGTATCCAGCACAAATCACAGTCAAACA 960
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
DB 961 GCTAGAGGTGTTCAGTGCACACGAGATCATTACCACCTTCACTCCCTTACTCTCAAATG 1020
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
DB 1021 TCTGAATTTGGAAGAACGAATCGCTCGTATTATTCCTTCTGTTATCGTTCACCACTTGG 1080
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
DB 1081 GTACCAGATTCAAGGCCAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAGGC 1140
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
DB 1141 CCGCAACTCGACCAAACTTTAAATAGACTCAAAATCTTCTTTGGTTAGTCAGCTGCTGA 1200
QY 381 ArgLysValGlyGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
DB 1201 CGAAAGTGTGGGAAGGATATGTTATCGAAGAAAGGCGATCTCTCGTTATGTTTGGC 1260
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
DB 1261 AAAGATTACCATCTGAACCTGTTAAATAATCTTGAAGCAAGTTATCAAAACAAAGAGAGT 1320
QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
DB 1321 GTTTCACACTTTAATCTGCTTAAAGAAAGAAATGTTGCTCTCTGTGACCAAGAAATTTAT 1380
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
DB 1381 GATAAAGCATATAATCTGTTAACTGAGGCTATAAAGCCCTTGTTCGNAATAGGGTCGT 1440
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
DB 1441 AATTCGTGATTCAGAGCTTAGACAAATATTAGAAGCGTTGAATGATGATGATGATGAT 1500
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
DB 1501 AAAGAAAATTTGATGATGATTTATTGGGCTTCTTAGCACCAATTTACCCATCCAGAGCGA 1560
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
DB 1561 CTTGGCAAAACCAAAATCTCAAATTTAGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
DB 1621 GCTGATAGGTATACACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
DB 1681 GAAGGAGATGATATGTAACGCTCATATGGGCCATAGTCACCTGGGATGGAAGATAGC 1740

```

```

QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaThrLysGluLysGlyIleLeuPro 580
DB 1741 CTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCTATACATAAGAAAAAGGTATCTCTACCT 1800

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIleTyr 600
DB 1801 CCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTATTATAC 1860

QY 601 AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis 620
DB 1861 AATCTGTGCAAGGGGAAACAGCAATTCACCTCGTTCGACTTCCATATATGTTGAGCAT 1920

QY 621 ThrValGluValLysAsnGlyAsnLeuIleProHisLysAspHisTyrHisAsnIle 640
DB 1921 ACAGTTGAGGTTAAAAACCGTAATTTGATTATTCCTCATAGGATCATTAACATAATATT 1980

QY 641 LysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp 660
DB 1981 AAATTTGCTTGGTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGAT 2040

QY 661 LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp 680
DB 2041 TTGTTTGGACGATTAAAGTACTACGTAGAACACCCCTGACGAACTCCACATTTCTAATGAT 2100

QY 681 GlyTyrGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700
DB 2101 GGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACACACAGTGAAGATCCAAAT 2160

QY 701 LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
DB 2161 AAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGCTGAGCCAGAGTCCCT 2220

QY 721 GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
DB 2221 CAAGTAGAGACTCAAAAAGTAGAAGCCCAACTCAAAAGAAAGCAGAAAGTTTGTCTTGCAGAA 2280

QY 741 ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
DB 2281 GTAACGGATTCTAGTCTGAAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 2340

QY 761 LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
DB 2341 TTGACTCTTCAAAATTATGATACATAGTATCATGCGAGAGCAGAGAAATTTACTTTCG 2400

QY 781 LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
DB 2401 TTGTTAAAGGAAGTAATCTCTCATCTGTAAGTAAGCAAAAAATAAAC 2448

RESULT 4
US-10-387-783-9
; Sequence 9, Application US/10387783
; Publication No. US2004005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 9
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: (1)-(2451)
; OTHER INFORMATION: n = a, c, t or g
US-10-387-783-9

Alignment Scores:
Pred. No.: 0 Length: 2451
Score: 4163.00 Matches: 796
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.95% Indels: 0
DB: 16 Gaps: 0

US-09-765-271-56 (1-796) x US-10-387-783-9 (1-2451)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20
DB 61 TCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGAAAAATAATCGTGTTCCTAT 120

QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 121 ATAGATGGAAACCAAGCGCGCAAAAAACGGAGAAATTTGACTCTCTGATCAGGTTAGCAAG 180

QY 41 ArgGluGlyIleAsnAlaGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
DB 181 CGTGAAGGAATCAATGCTGAGCAAAATCGTCAAGATAACAGACCAAGGCTATGTCACT 240

QY 61 SerHisGlyAspHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB 241 TCACATGGCGACCCATCTATTATACATGTTAGGTTTCTTNTGACGCTATCATCAGT 300

QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
DB 301 GAAGAAATTACTCATGAAGATCCAACTATAAGTAAAAAGATGAGGATATTGTTAATGAG 360

QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAspAla 120
DB 361 GTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGGATGCT 420

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
DB 421 GCCACGCGGATTAACGTCCTGTAACAAAGAGAAATCAATCGCAAAAAACAAGACATAGT 480

QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 481 CAACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGCTGTGCTTGGCCTTGGCAGCTTGC 540

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 541 GGACGCTATACATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGGATGACT 600

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
DB 601 GGTGATGCTTATATCTGTTCTCTCATGAGATCATTAACATTAACATTTCTTAAGATGAGTTA 660

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 661 TCAGCTAGCGAGTTGCTGCTGCGAGAGCCTTCTATCTGCTGAGGAAATCTGTCAAAAT 720

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
DB 721 TCAAGAACCTTATCGCGCAAAAAATAGCGATAACACTTCAAGAAACAACTGGGTACCTTCT 780

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
DB 781 GTAAGCAATCCAGAACTACAAATATAACAAAGCAACCAACAGCAACTAAACAGTCAA 840

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer 280
DB 841 GCAAGTCAAAAGTAATGACATTTGATGCTCTTGAACACAGCTCTCAAAACGCTTTGAGT 900

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300

```



901	DB	CAAACGACATGTAGAATCTGATGCCCTTGCTCTTTGATCAGACACAAATCACAAGTCGAACA	960
301	QY	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
961	DB	GCTAGAGGTGTTCAGTGCACACGAGATCAATACCACTTCATCCCTTACTCTCAATG	1020
321	QY	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
1021	DB	TCGTAAATTTGGAGAAGCAATCGCTCGTATATTCCCTTCCTGTTATCGTTCAACCATGG	1080
341	QY	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
1081	DB	GTACCAGATTTCAGGCCAGAAACAACAACTCCACACCGACTCCGGAACCTAGTCCAGGC	1140
361	QY	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
1141	DB	CGCAACCTGCACCAAAATCTTAAATAGACTCAAAATTTCTTTGGTTAGTCAGCTGGTA	1200
381	QY	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
1201	DB	CGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGCACTCTCTGTTATGCTTTGCG	1260
401	QY	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
1261	DB	AAAGATTTACCATCTGAACCTGTATAAAATCTCTGAAAGCAAGTTATCAAAACAAGAGAGT	1320
421	QY	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
1321	DB	GTTTTCACACACTTTAACTGCTAAAAAAGAAAAATGTTGCTCTCGTGACCAAGAAATTTTAT	1380
441	QY	AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
1381	DB	GATAAAGCATATATCTCTGTTAACTGAGGCTCATAAAGCCCTTGTTTGNAAATAAGGTCGT	1440
461	QY	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
1441	DB	AAATCTGATTTCCNAGCCTTAGACAAATATTAGAAGCGCTTGAATGATCAATCGACTAAT	1500
481	QY	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
1501	DB	AAAGAAAAATTGGTAGATGATTTATTGGCATTCCTAGCACCAATATCCCATCCAGAGCGA	1560
501	QY	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520
1561	DB	CTTTGGCAACCAAAATCTCAAATTTGAGTATCTGGAAGACGAAGTTCTGTTATCTCAATTA	1620
521	QY	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp	540
1621	DB	GCTGATAAGTATACACGTCAGATGGTTACATTTTTCATGAAATCATGATATATCATGCTAT	1680
541	QY	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyIleAspSer	560
1681	DB	GAAGGAGATGCATATGTAACGCTCATATGGGCCATAGTCATCTGGATTGGAAAAAGATAGC	1740
561	QY	LeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro	580
1741	DB	CTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATCTAAAGAAAAAAGGTATCCTACCT	1800
581	QY	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr	600
1801	DB	CCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAGTGCACGACGACTATTATC	1860
601	QY	AsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
1861	DB	AAATCGTGTGAAGGGGAAAAAGCAATTCACCTCGTTCCACTTATATGTTGGTGGAGCAT	1920
621	QY	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
1921	DB	ACAGTTGAGGTTAAAAACGGTAATTTGATATTTCCTCATAAGGATCATTACCAATAATT	1980
641	QY	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
1981	DB	AAATTTGCTTGGTTTGATGATCACACATCAAAAGCTCCAAATGGCTATACCTTCCGAAGAT	2040

QY	661	LeuPheAlaThrIleLysGlyTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	2041	TTGTTTGCGACGATTAAAGTACTACGTAGAACACCCTGACGAACTGCATTTCTTAATGAT	2100
QY	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700
Db	2101	GGATGGGGCAATGCCAGTGAGCATGTGTAGCGAAGAAGACCACTAGTAGGAATCCCAAT	2160
QY	701	LysAsnPheLysAlaAspGluProValGluThrProAlaGluProGluValPro	720
Db	2161	AAGAACTTCGAAGCGGATGAGAGCCAGTAGAGGAACACCTGCTGAGCAGAGTCCCT	2220
QY	721	GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys	740
Db	2221	CAAAGTAGAGACTGAAAAAGTAGAAGCCCACCTCAAGAGCAGAAAGTTTTGCTTGCAAA	2280
QY	741	ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn	760
Db	2281	GTAACGGATTCTAGTCTGAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAAATAAT	2340
QY	761	LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla	780
Db	2341	TTGACTCTTCAAAATTATGGATAACAATAGTATCATGCGAAGCAGAAAAATTACTTGGC	2400
QY	781	LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn	796
Db	2401	TGTGTAAGAAGGAAGTAATCTTCTATCTGTAAGTAAGCAAGAAAAATAAAC	2448
<b>RESULT 5</b>			
US-10-158-844-94			
; Sequence 94, Application US/10158844			
; Publication No. US20040029118A1			
; GENERAL INFORMATION:			
; APPLICANT: Kunsch et al.			
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences			
; NUMBER OF SEQUENCES: 391			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Human Genome Sciences, Inc.			
; STREET: 9410 Key West Avenue			
; CITY: Rockville			
; STATE: Maryland			
; COUNTRY: USA			
; ZIP: 20850			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: CD-R			
; COMPUTER: Dell Latitude Pentium 3			
; OPERATING SYSTEM: Windows 98			
; SOFTWARE: ASCII Text			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/10/158,844			
; FILING DATE: 03-Jun-2002			
; CLASSIFICATION: <Unknown>			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/961,527			
; FILING DATE: 1997-10-30			
; APPLICATION NUMBER: US 60/029,960			
; FILING DATE: 1996-10-31			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Hyman, Mark J.			
; REGISTRATION NUMBER: 46,789			
; REFERENCE/DOCKET NUMBER: PB340PID1			
; INFORMATION FOR SEQ ID NO: 94:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 8195 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:			
US-10-158-844-94			
Alignment Scores:			
Pred. No.:                      0                      Length:                      8195			

Score:	4163.00	Matches:	795
Percent Similarity:	99.87%	Conservative:	0
Best Local Similarity:	99.87%	Mismatches:	1
Query Match:	99.95%	Indels:	0
DB:	13	Gaps:	0
US-09-765-271-56 (1-796) x US-10-158-844-94 (1-8195)			
Qy	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVallysGluAenAsnArgValSerTyr	20
Db	3054	TCITACGAGTTGGACCTGATCAAGCTAGAAGCGTTAAGGAAATAAATCGTGTTCCTAT	3113
Qy	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
Db	3114	ATAGATGMAAACCAAGCGACGCAAAACCGGAGAAATTTGACTCCCTGATGAGGTAGCAAG	3173
Qy	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
Db	3174	CGTGAAGGAATCAATGCTGAGCAATCGTCACAGATAACAGACCAAGGCTATGCTACT	3233
Qy	61	SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
Db	3234	TCACATGGCGACCACTATCATTTATTAATGTAAGGTTCCCTATGACGCTATCATCAGT	3293
Qy	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
Db	3294	GAAGAATTTACTCATGAAAGATCCAACTATAAGCTAAAGATGAGGATATTTGTTAATGAG	3353
Qy	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
Db	3354	GTCAGGGTGGATATGTTATCAAGGTAGTGAATAATACTATGTTTACCTTAAGGATGCT	3413
Qy	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer	140
Db	3414	GCCACGCGGATAACCTCCGTA CAAGAAGAGAAATCAATCGCAAAACCAAGAGCATAGT	3473
Qy	141	GlnHisArgGluGlyThrProArgAsnAspGlyValAlaValAlaAlaArgSerGln	160
Db	3474	CAACATCGTGAAGGTGGAACTCCAGAAACGATGGTGGTCTTCCCTGGCACGTTGCGAA	3533
Qy	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
Db	3534	GGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTCATATCATAGAGGATACT	3593
Qy	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
Db	3594	GGTGAIGCTTATATCGTCTCTCATGGAGATCAATCACTTACATTCCTAAGAATGAGTTA	3653
Qy	201	SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
Db	3654	TCAGCTAGCGAGTTGGCTGCTGCAAGAGCCCTTCTATCTGGTCGAGGAATCTGTCBAAT	3713
Qy	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
Db	3714	TCAGAACCTATCCCGCAAAAATAGCGATAACACTTCAAGAACAACTGGGTACCTCTCT	3773
Qy	241	ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
Db	3774	GTAAGCAATCCAGAACTACAAATACTAACACAGCAACACACACACACTAACAGTCAA	3833
Qy	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
Db	3834	GCAAGTCAAGATATGATTTGATAGTCTTCTTGAAACAGCTCTACAACTGCCCTTGAGT	3893
Qy	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
Db	3894	CAACGACATGTAGAATCTGATGCGCTTCTTCTTGCATCCAGCACAAATCACAGTCGAACA	3953
Qy	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
Db	3954	GCTAGAGGTGTTCAGTGCCACCGAGATCATACCACTTCACTTCTTCTTCTTCAATG	4013
Qy	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340

Db	4014	TCTGAATTTGAAGAACGAATCGCTCGTATTATTTCCCTTCTGTTATCGTTCAACCAATGG	4073
Qy	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
Db	4074	GTACCAATTTCAAGGCCAGAACCAACCAAGTCCACCAACCGACTCCGGAAACCTAGTCCAGGC	4133
Qy	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
Db	4134	CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTCTTTGGTTAGTCAGCTGGTA	4193
Qy	381	ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
Db	4194	CGAAAGTTGGGGAAGGATATGATTCGAGAAAGGGCACTCTCCGTTATGTCTTTGCG	4253
Qy	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
Db	4254	AAAGATTTTACCATCTGAACCTGTTAAATCTTGAAGCAAGATTATCAAAACAAGAGAT	4313
Qy	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
Db	4314	GTTCACACACTTTAACTGTCTAAAGAAAGAAATGTTCTCTCTGTCGACCAAGAAATTTAT	4373
Qy	441	AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
Db	4374	GATAAAGCATATAATCTGTCTTAACTGAGGCTCATAAAGCTTGTGTTGAAATAAGGTCGT	4433
Qy	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
Db	4434	AAATCTGATTTCCAGACCTTAGACAAATTTATTAGAACGCTTGAATGATGAATCGACTAAT	4493
Qy	481	LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg	500
Db	4494	AAAGAAAAATTTGGTATGATTTATTGGCAATTTCTTAGCACCAATTACCCATCCAGACCGA	4553
Qy	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520
Db	4554	CTTGGCAAACTTCTCAAAATTTGATATCTAGAGACGAAGTTCGTATGCTCAATTA	4613
Qy	521	AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp	540
Db	4614	GCTGATAGTATACACGTCAGATGTTTACATTTTGTGATGAACATGATATAATCAGTGAT	4673
Qy	541	GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer	560
Db	4674	GAAGGAGATGCATATGATTAACGCTCATATGGCCATAGTCACGGAATGGAAAGATAGC	4733
Qy	561	LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro	580
Db	4734	CTTCTGTAAGGAAAGAAAGTTGCGCTCAAGCTTACTAAGCTTAAAGAAAAGGTATCTACCT	4793
Qy	581	ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr	600
Db	4794	CCATCTCCAGACCGCAGATGTTAAAGCAAACTCCAACTGGAGATAGTCAGACAGCTATTATC	4853
Qy	601	AsnArgValLysGlyLysArgIleProLeuValArgLeuProTyrMetValGluHis	620
Db	4854	AATCGTGTGAAAGGGGAAACCAATTTCCACTCGTTCGACTTCCATATATGTTGAGCAT	4913
Qy	621	ThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsnIle	640
Db	4914	ACAGTTGAGGTTTAAACCGGTAAATTTGATTTCTCTCATAGGATCATTAACCAATATTT	4973
Qy	641	LysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAsp	660
Db	4974	AAATTTGCTGTTTGTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGAT	5033
Qy	661	LeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAsp	680
Db	5034	TTGTTGCGACGATTAAGTACTAGTAGAACCCCTCGACACCGTCCACATTTCTAATGAT	5093
Qy	681	GlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn	700

```

Db      5094 GGATGGGGCAATGCCAGTGAGCATGTGTTAGGCAAGAAAGACCAACAGTGAAGATCCAAAT 5153
Qy      701  LysAsnPheLysAlaAspGluProValGluGluThrProAlaGluProGluValPro 720
Db      5154 MAGAACTTCAAGCGGATGAAGAGCCAGTAGAGGAAACACCTGCTGAGCCAGAGTCCCT 5213
Qy      721  GlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLys 740
Db      5214 CAAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAGCAGAGATTGCTTGGGAAA 5273
Qy      741  ValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsn 760
Db      5274 GTAACGGATTCTAGTCTGAAAGCCCAATGCAACAGAAACTCTAGCTGGTTTACGAAATAAT 5333
Qy      761  LeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAla 780
Db      5334 TTGACTCTTCAAAATTATCGATAACATAGTATCATGGCAGAGCAGAGAAATTAATTGGC 5393
Qy      781  LeuLeuLysGlySerAsnProSerSerValSerLysGluLysIleAsn 796
Db      5394 TTGTTAAAGGAAGTAATCTTCTCATCTGTAGTAAGGAAAAATAAAC 5441

RESULT 6
; Sequence 3, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-884-465A-3

Alignment Scores:
Pred. No.: 2,37e-282 Length: 2523
Score: 3218.00 Matches: 616
Percent Similarity: 82.95% Conservative: 65
Best Local Similarity: 75.03% Mismatches: 102
Query Match: 77.26% Indels: 38
DB: 10 Gaps: 3

US-09-765-271-56 (1-796) x US-09-884-465A-3 (1-2523)

Qy      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db      61  GCTTATGAAGTACTAGTTTGCATCAAGCTCAACTGTAAGAGAAATAATCGTGTTCCTAT 120
Qy      21  IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db      121  ATAGATGGAAAAACAAGCGCAACAAAAACGAGAAATTTGACTCTCTGATGGGTAGCAAG 180
Qy      41  ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db      181  CGTGAAGGAATCAACGCCCAACAAATCGTCATCAAGATTACGGATCAAGGTTATGTGACC 240
Qy      61  SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db      241  TCTCATGGAGACCATTTATCTACTATAATGGCAAGGTCCTTATGATGCCATCATCAGT 300

```

```

Qy      81  GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db      301  GAAGAGCTCTCATGAAAGATCCGAATTATCAGTTGAAGATTTCAGACATTGTCAATGAA 360
Qy      101  ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db      361  ATCAAGGGTGGTTATGTCTAATAGGTAAACGGTAATACTATGTATTACCTTAAGGATGCA 420
Qy      121  AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db      421  GCTCATCGGATAATGTCCGTACAAAGAGAAATCAATCGCAAAACACAGAACATAGT 480
Qy      141  GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db      481  CAGCATCTGTGAAGGAGGACTTCAGCAAAACGATGTCGGTAGCCTTTGCGACGTTCCACAG 540
Qy      161  GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db      541  GGACGCTACACACAGATGATGGTTATATCTCAATGATCATCTGATATCATCGAAGATAGC 600
Qy      181  GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db      601  GGCGATGCTATATCGTTCTCTCATGGAGATCATACCATTTACATTCTCTTAAGATGAGTTA 660
Qy      201  SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db      661  TCAGCTAGCGAGTTGGCTGCTGCAGAGCCCTCTCTATCTGGTCGGGAAAAATCTGTCAAA 720
Qy      221  SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer 240
Db      721  TTAAGAACCTTATCCCGCACAAAATAGCGATAACACTCCAGAAACAAACTGGGTACTTCT 780
Qy      241  ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db      781  GTAAGCAATCCAGAACTACAAATACTAACACAGCAACACACACACACTTAACGTCGTA 840
Qy      261  AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuLysGlnLeuLysLeuProLeuSer 280
Db      841  GCAAGTCAAAAGTAATGACATTGATGATCTCTTGAACACAGCTCTACAAACTGSCCTTTGAGT 900
Qy      281  GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db      901  CAACGCCATGTAGATCTGATGGCTTTATTTTCGACCCAGCGCAAAATCACAAAGTCGAACC 960
Qy      301  AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db      961  GCCAGAGGTGATGCTGTCCTCATGTGTAACCATTAACCACTTTATCCCTTTAAGCAAAATG 1020
Qy      321  SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTyr 340
Db      1021  TCTGAATTTGAAAAACGAATTGCTCGTATTATTTCCTTCCTGTTATCGTTTCAAAACCATTTGG 1080
Qy      341  ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db      1081  GTACACAGATTCAGACACAGAAACCAAGTCCAAACCGACTCCAGAACCTAGTCCCAAGT 1140
Qy      361  ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer 374
Db      1141  CCGCAACCTGCACCAATCTCAACCCAGCTCCCAAGCAATCCAATTGAT-----GAGAAA 1194
Qy      375  LeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIle 394
Db      1195  TTGGTCAAAAGAGCTGTTCGAAAAAGTAGCGATGTTATGTCTTTGAGGAGATGAGATT 1254
Qy      395  SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys 414
Db      1255  TCTCGTTATTCGCCCAAGAAATCTTTCAGCAGAAACACAGCAGCAGCATTTATGATGACAAA 1314
Qy      415  LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAlaPro 434
Db      1315  CTGGCCACAGCAGAAAGTTTATCTCATAGCTAGAGCTTAAGAAACACTGACCTCCCATCT 1374
Qy      435  ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu 454

```

Db	1375	AGTCATCGAAGATTTTCAATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTA	1434	Db	2455	ACCCAGGACAAACATACTATTATGGCAGAAAGCTGAAAACACTATTGGCTTTTATTAAAGGAG	2514
Qy	455	Phe**AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeu	474	Qy	785	Ser 785	
Db	1435	CTTGATATAAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTGGAACGACTC	1494	Db	2515	AGT 2517	
Qy	475	AsnAspGluSerThrAsnLysGluLysLeuValAspLeuLeuAlaPheLeuAlapro	494	RESULT 7			
Db	1495	AAGGATGCTCAAGTGATAAAGTCAAGTTAGTGATGATATCTTGCCCTTTTAGCTCCG	1554	US-09-884-465A-4			
Qy	495	IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluLysThrGluAspGlu	514	; Sequence 4, Application US/09884465A			
Db	1555	ATTGCTATCAGAACAGTTTAGAAGAACCAATAGCGCAATATCTCACTGATGATGAG	1614	; Publication No. US20030077293A1			
Qy	515	ValArgIleAlaGlnLeuAlaAspLysThrThrSerAspGlyTyrIlePheAspGlu	534	; GENERAL INFORMATION:			
Db	1615	ATTCAAGTAGCAAGTTGCGAGGCAAGTACACACAGAGACGGTTATATCTTTGATCCT	1674	; APPLICANT: Shire Biochem, Inc.			
Qy	535	HisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHis	554	; APPLICANT: Hamel, Josee			
Db	1675	COTGATATAACCAAGTGATGAGGGGATGCTATGTAATGTAATCCATATGACCCATGCCAC	1734	; APPLICANT: Brodeur, Bernard			
Qy	555	TrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLys	574	; APPLICANT: Martin, Denis			
Db	1735	TGGATTAATAAAGATAGTTGTCGAAGCTGAGAGCGCGACGCCAGCTTATGCTAAA	1794	; APPLICANT: Charland, Nathalie			
Qy	575	GluLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGlyAsp	594	; APPLICANT: Ouellet, Catherine			
Db	1795	GAGAAAGGTTTGACCCCTCTTCGACAGACCATCAGGATTCAGAGAAATACTGAGGCAAAA	1854	; TITLE OF INVENTION: Streptococcus Antigens			
Qy	595	SerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArgLeu	614	; FILE REFERENCE: 055190-0044			
Db	1855	GGAGCAGAGAGCTATCTACAAACCGCGTAAAGAGCGTAAAGAGGTGCCATCTGATCGTAG	1914	; CURRENT APPLICATION NUMBER: US/09/884,465A			
Qy	615	ProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHisLys	634	; CURRENT FILING DATE: 2001-06-20			
Db	1915	CTTTACAATCTTCAATATACTGTAGAGTCAAACCGGTAGTTTAATCATACCTCATTAT	1974	; PRIOR APPLICATION NUMBER: 60/212,683			
Qy	635	AspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsn	654	; PRIOR FILING DATE: 2000-06-20			
Db	1975	GACCATTAACATAACATCAAAATTTGAGTGGTTTGACGAAGGCGCTTTATGAGGCACCTAAG	2034	; NUMBER OF SEQ ID NOS: 384			
Qy	655	GlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGlu	674	; SOFTWARE: Patentin version 3.1			
Db	2035	GGGTATACCTCTTGAGGATCTTTGGCGACTGTCAAGTACTAGTCTAGCGACCATGTTCAAAAGAACAAAT	2094	; SEQ ID NO 4			
Qy	675	ArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAsp	694	; TYPE: DNA			
Db	2095	CGTCCGCTTCAGATTAATGTTTGGTAAAGCTAGCGACCATGTTCAAAAGAACAAAT	2154	; ORGANISM: Streptococcus pneumoniae			
Qy	695	HisSerGluAspProAsnLysAsnPheLysAlaAsp	706	; US-09-884-465A-4			
Db	2155	GGTCAAGCTGATACCAATCAACCGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAA	2214	Alignment Scores:			
Qy	706	-----	706	Pred. No.:	2,53e-282	Length:	2647
Db	2215	CTTGAGGAGAAACCCCTCGAGAGAGAAACCAACAGGAGGAGAAACCTCAGAGTCTCCAAAA	2274	Score:	3218.00	Matches:	616
Qy	707	-----GluGluProValGluGluThrProAlaGluProGluValProGluValGluThr	724	Percent Similarity:	82.95%	Conservative:	65
Db	2275	CCACAGAGAGAACCAAGAGAGAGATCACCAGAGGATCAGAGAACTCAGGTCGAGACT	2334	Best Local Similarity:	75.03%	Mismatches:	102
Qy	725	GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer	744	Query Match:	77.28%	Indels:	38
Db	2335	GAAGAGGTTGAAGAAAAAAGCTGAGAGAGGCTGAGAGATTCTTGAAAAATCCAGAGTCCA	2394	DB:	10	Gaps:	3
Qy	745	SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGln	764	US-09-765-271-56 (1-796) x US-09-884-465A-4 (1-2647)			
Db	2395	ATTATCAGTCCCAATGCGCAAGAGACTCTCAGAGATTAAATATTTACTATTGGC	2454	Qy	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr	20
Qy	765	IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGly	784	Db	105	GCTTATGAATAGTGGTTTGCATCAAGCTCAATCTGTATTAAGAAATATCGTGTTCCTAT	164

QY	161	GlyArgTyrThrThraspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
DB	585	GGACGCTACACACAGATGATGTTATATCTTCAATGCATCTGATATCATCGAAGATACG	644
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
DB	645	GCGGATGCCCTATCGTTCTCTCATGGAGATCATACCATTAATTCCTAAGAATGAGTTA	704
QY	201	SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
DB	705	TCAGCTAGCGAGTTCGGCTGCTGCAGAGCCTTCTTATCTGGTCGGAAAAATCTGTCAAA	764
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer	240
DB	765	TTAAGAACCTATCGCCGCAAAAATAGCATAACACTCCAAGAACAACCTGGGTACCTTCT	824
QY	241	ValSerAsnProGlyThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln	260
DB	825	GTAACCAATCCAGGAACCTACAAATACTAACCAAGCAACAACAGCAACACTAACAGTCAA	884
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
DB	885	GCAAGTCAAAGTAATGACATTTGATAGTCTCTTGAACAGCTCTACAACTGCCCTTGAGT	944
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
DB	945	CAACGCCATGTAGAACTTCGATGGCTATTATTTTCAGCCAGCGCAATCACAAGTCGAACC	1004
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
DB	1005	GCCAGAGTGTAGCTGTCCCTCATGGTAACCATTAACCACTTATTCCTTATGAACAATG	1064
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
DB	1065	TCTCAATTGGAAAAACGAATTGCTCGTATATTCCCTCTCGTTATCGTTCAACCATTGG	1124
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
DB	1125	GTACCAAGATTCAAGACCAGAAGAACCAAGTCCACAACCGACTCCAGAACCTAGTCCAAGT	1184
QY	361	ProGlnProAlaProAsnLeuLys-----IleAspSerAsnSerSer	374
DB	1185	CCGCAACTGCACCAAAATCTCAACCAAGCTCCAAGCAATCCAATTGAT-----GAGAA	1238
QY	375	LeuValSerGlnLeuValArgLysValGlyGlyGlyTyrValPheGluGluLysGlyIle	394
DB	1239	TTGGTCAAGAAGCTGTTTCGAAAAGTAGGCGATGTTATGTTCTTTGAGGAGAATGGAGTT	1298
QY	395	SerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLys	414
DB	1299	TCTCGTTATATCCAGCCCAAGAATCTTTCAGCAGAAACAGCAGCAGCAATTGATAGCAA	1358
QY	415	LeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluLeuValAlaPro	434
DB	1359	CTGCCAAGCAGGAAGTTTATCTATAAGCTAGAGCTAAGAAAACTGACCTCCCATCT	1418
QY	435	ArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeu	454
DB	1419	AGTGATCGAGATTTTACAATAAGGCTTATGACTTACTAGCAAGAATTCACCAAGATTTA	1478
QY	455	Phe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuArgLeu	474
DB	1479	CTTGATAATAAAGGTCGACAAAGTTGATTTTGGGCTTTGGATAACCTGTTTGAACGACTC	1538
QY	475	AsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaPro	494
DB	1539	AAGGATGCTCAAGTGAATAAGTCAAGTATAGTGGATGATATTTCTTCCCTTCTTAGCTCG	1598
QY	495	IleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGlu	514
DB	1599	ATTGCTCATCCAGACGTTTAGAAAAACCAATAGCCAAATTAACCTACATGATGATGAG	1658
QY	515	ValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGlu	534

## RESULT 8

```

RES011 8
US-09-769-787-206
? Sequence 206, Application US/09769787
? Publication No. US20030091577A1
? GENERAL INFORMATION:
? APPLICANT: Microbial Technics Limited
? APPLICANT: Gilbert, Christophe FG
? APPLICANT: Hansbro, Philip M
? TITLE OF INVENTION: proteins
? FILE REFERENCE: PWC/P21129W0

```

Fri Oct 1 15:35:52 2004

;; CURRENT APPLICATION NUMBER: US/09/769,787  
;; CURRENT FILING DATE: 2001-01-26  
;; PRIOR APPLICATION NUMBER: GB 9816337.1  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: US 60/125164  
;; PRIOR FILING DATE: 1999-03-19  
;; NUMBER OF SEQ ID NOS: 388  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 206  
;; LENGTH: 2481  
;; TYPE: DNA  
;; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-206

Alignment Scores:  
Pred. No.: 6,58e-244 Length: 2481  
Score: 2795.00 Matches: 549  
Percent Similarity: 76.98% Conservativity: 83  
Best Local Similarity: 66.87% Mismatches: 137  
Query Match: 67.11% Indels: 52  
DB: 10 Gaps: 8

US-09-765-271-56 (1-796) x US-09-769-787-206 (1-2481)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19  
DB 61 TCCTATGAGCTGGTGGTCGTCACCAAGCTGGTCAGGATAGAAAGAGTCTATCGAGTTGCT 120  
QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
DB 121 TATATAGATGGTGATCAGCTGGTCAAAGGCGAGAAACTTCACACCAGATGAAGTCAGT 180  
QY 40 LysArgGluGlyLeuAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
DB 181 AAGAGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGATCAAGGTTATG 240  
QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79  
DB 241 ACCTCTCATGGAGACCATATCATTAATTAATGGCAAGGTCCTTATGATGCCATATC 300  
QY 80 SerGluGluLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
DB 301 AGTGAGAGCTCTCATGAAGATCGGAATTATCAGTTGAAGGATTACAGACATTTGCAAT 360  
QY 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysAsp 119  
DB 361 GAAATCAAGGGTGGTTATGTGTATCATCAAGGTAGACGGAAATATCTATGTTACCTTAAGGAT 420  
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
DB 421 GCAGCTCATCGGATATATTCGACAAAGAGAGATTAACGTCAGAGCAGGAACGC 480  
QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
DB 481 AGTCATAATCAC-----GGGTCAAGGAGCTAACGATCATGCTAGTGTGAGCCAGAGCC 534  
QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
DB 535 CAAGGAGCGCTATACACGGATGATGGGTATATCTTCAATGCATCTGATATCATITAGGAC 594  
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
DB 595 ACGGGTGATGCTTATATCGTTCCTCAGCGGACCATTAACCATTAATTCCTAAGAATGAG 654  
QY 200 LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
DB 655 TTATCAGCTAGCAGGTTAGCTGCTGCAGAGCCTATTGGAATGG----- 699  
QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
DB 700 -----AAGCAGGGATCTCGCTCTTCAAGTCTTAGTTATATGCA 741  
QY 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259

DB 742 AATCCAGCTCAACCAAGATTGTTCAGACACCAAAATCTGACTGTCTCACTCCAACTTATCAT 801  
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279  
DB 802 CAA--AATCAAGGGGAAACATTTCAAGCCTTTACGTGAATGTATGCTAAACCCCTTA 858  
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
DB 859 TCAGAACGCCATGTGAATCTGATGGCTTATTTTCGACCCAGCGCAATCACAAGTCGA 918  
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
DB 919 ACCGCCAGAGGTAGCTGTCCCTCATGGTAACCAATTACCACTTTATCCCTTATGAAACA 978  
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
DB 979 ATGCTGAATTTGGAAAAAGCAATTCCTGCTATTTATCCCTTCGTTATCGTTCAACCAT 1038  
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
DB 1039 TGGGTACAGATTCAAGACCAACAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1098  
QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
DB 1099 AGTCGCAACCTGCACCAAAATCTCAACCCAGCTCCCAAGCAATCCAAATGAT-----GAG 1152  
QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
DB 1153 AAAATGGTCAAGAAAGCTGTTCCAAAAGTAGCGGATGTTATGTCTTGGAGAGAATGCA 1212  
QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
DB 1213 GTTTCCTGTTATATCCAGCCCAAGGATCTTTCAGCAGAAACAGCAGCAGCATGTATAGC 1272  
QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
DB 1273 AAACTGGCCCAAGCAGAAAGTTTATCTCATAGCTAGGAGCTAAGAAACTACCTCCCA 1332  
QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
DB 1333 TCTAGTATCGAGATTTTACAAAGGCTTATAGCTTACTAGCAAGAATTCACCAAGAT 1392  
QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuGluArg 473  
DB 1393 TTACTTTGATAATAAAGGTCGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTGGAACGA 1452  
QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
DB 1453 CTCAAGGATGTCCCAAGTCAAGTCAAGTTAGTGTGATGATATTCTTGCCTTCTTAGCT 1512  
QY 494 ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
DB 1513 CCGATTTCGTCACTCCAGAACGTTTAGGAAAAACCAATGCGCAAAATTACCTACACTGAT 1572  
QY 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
DB 1573 GAGATTCAAGTAGCCCAAGTTGGCAGGCAAGTACACAAACAGAACGCTTATATCTTTGAT 1632  
QY 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
DB 1633 CCTCGTGATATAACCGATGATGAGGGGATGCTTATGTAATCTCACAATAGCCCATAGC 1692  
QY 554 HistripleGlyLysAspSerLeuSerAspLysGluLysValAlaGlnAlaTyrThr 573  
DB 1693 CACTGGATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGCGAGCCAGGCTTATGCT 1752  
QY 574 LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly 593  
DB 1753 AAAGAGAAAGGTTGACCCCTCTCTGACAGACCATCAGGATTCAGGAATTACTGAGGCA 1812  
QY 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg 613



Db 1813 AAAGGAGCAGAAGCTATCTTACAACCGCGTGAAAGCAGCTAAGAAGGTGCCACTTTGTCGT 1872  
 QY 614 LeuProTyMetValGluHisThrValGluValLysAsnGlyAsnLeuIleileProHis 633  
 Db 1873 ATGCCTTCAACTCTTCATATACTGTAGAGTCAAAAACGGTAGTTTAATCATACCTCAT 1932  
 QY 634 LysAspHisTyThrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlapro 653  
 Db 1933 TATGACCATTACCAAPAACATCAAATTTGAGTGGTTTGACGAAGCCGCTTTATGAGCACCT 1992  
 QY 654 AsnGlyTyThrLeuGluAspLeuPheAlaThrIleLysTyThrValGluHisProAsp 673  
 Db 1993 AAGGGGTATCTCTTGAGATCTTTTGGCGACTCTCAAGTACTATGTCGAACATCCAAAC 2052  
 QY 674 GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLys 693  
 Db 2053 GAACGTCCGCATTCTTGAGATCTTTTGGCGACTCTCAAGTACTATGTCGAACATCCAAAC 2112  
 QY 694 AspHisSerGluAspProAsnLysAsnPhelysAlaAspGluGlu----- 708  
 Db 2113 AATGGTCAAGCTGATACCAATCAAACCGMAAAACCAACGCGAGGAGAAACCTCAGACAGAA 2172  
 QY 709 --ProValGluGluThr----- 713  
 Db 2173 AAACCTGAGGAAGAAACCCCTCGAGAGAGAGAAACCGCAAAACGAGAAACAGAGTCTCCA 2232  
 QY 714 -----ProAlaGluProGluValProGlnValGluThr 724  
 Db 2233 AAACCAACAGAGAACCCAGAGAATCACCAGAGGATCAGAGAACCTCAGGTCGAGACT 2292  
 QY 725 GluLysValGluAlaGlnLeuLysGluAlaGluValLeuLeuAlaLysValThrAspSer 744  
 Db 2293 GAAAGGTTGAAGAAAAAAGCTCAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCA 2352  
 QY 745 SerLeuLysAlaAsnAlaThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGln 764  
 Db 2353 ATTATCAAGTCCATGTCAAAGAGACTCTCAGAGGATTAATAAATAATTTACTATTGGC 2412  
 QY 765 IleMetAspAsnAsnSerIleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGly 784  
 Db 2413 ACCCAGGACAACTACTATTATGGCAGAAGCTGAAAAACTTTGGCTTTTATAAAGGAG 2472  
 QY 785 Ser 785  
 Db 2473 AGT 2475  
 RESULT 9  
 US-10-412-862-11  
 ; Sequence 11, Application US/10412862  
 ; Publication No. US20040052781A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; PNEUMONIAE Polypeptides Having Selected Structural  
 ; MOTIFS  
 ; FILE REFERENCE: 469201-685  
 ; CURRENT APPLICATION NUMBER: US/10/412,862  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 11  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-862-11

Alignment Scores:  
Pred. No.: 4.01e-243 Length: 2531

Score:	2786.50	Matches:	544
Percent Similarity:	77.06%	Conservative:	84
Best Local Similarity:	66.75%	Mismatches:	140
Query Match:	66.90%	Indels:	47
DB:	13	Gaps:	7

US-09-765-271-56 (1-796) x US-10-412-862-11 (1-2531)

QY	1	SerTyrGluLeuGlyLeuTyrGlnAla---	ArgThrValIysGluAsnAenArgValSer	19
DB	61	TCCTATGACGTGGACGTTACCAAGCTGTCAGGATAAGAAAGAGTCTTAATCGAGTTGCT	120	
QY	20	TyrIleAspGlyLysGlnAlaThrGlnIysThrGluAsnLeuThrProAspGluValSer	39	
DB	121	TATATAGATGGTCATCAGGCTGGTCAAAAGGCAGAAAACTTGACACCAGATGAAGTCAGT	180	
QY	40	LysArgGluGlyLeuAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal	59	
DB	181	AAGAGGAGGGATCAACGCCAAACAAATGTGTATCAAGATTACGATCAAGGTTATGTG	240	
QY	60	ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle	79	
DB	241	ACCTCTCTCGAGACCATATCATTTACTATATTAATGCAAGGTTCCCTTATGTCGCAATC	300	
QY	80	SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn	99	
DB	301	AGTGAAGAGCTCCTCATGAAGATCCGAATATCATGTTGAAGATTTCAGACATTGTCAT	360	
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp	119	
DB	361	GAATCAAGGGTGGTTATGTCATTAAAGGTAAACGGTAATACTATGTTTACCTTAAGGAT	420	
QY	120	AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis	139	
DB	421	GCRCCTCATCGGATAATATTCGCACAAAGAGAGATTAAACGTCAGAAGCAGAAACGC	480	
QY	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159	
DB	481	AGTCATATCATAACTCA-----AGACGAGATAATGCTGTTGCTCCAGCCAGAGCC	531	
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp	179	
DB	532	CAAGAGCTTATAACAAGCATGATGGGTATATCTTCAATGATCTGTATCATTCAGGAC	591	
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu	199	
DB	592	ACGGTGTGCTTATATCGTTCCTCACGGCGACCATTAACCATTTACATTCCTTAAGAATGAG	651	
QY	200	LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219	
DB	652	TTATCACTACGAGTTAGTGTCTGCAGAAGCCTATTGGAATGG-----	696	
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro	239	
DB	697	-----AAGCAGGGATCTCGTCTTCTTCAAGTCTTAGTTAATATGCA	738	
QY	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259	
DB	739	AATCCAGCTCAACCAAGATTGTCAGAAACCAACTCTGACTGTCATCTCCAACTTATCAT	798	
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279	
DB	799	CAA--AATCAAGGGGAAAACATTCAAGCCTTTTACGTGAATTGTATGCTAAACCCCTTA	855	
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299	
DB	856	TCGAAGCCCATGTGGAATCTGATGGCCCTTAITTTTCGACCCAGCGCAATACACAAGTCGA	915	
QY	300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319	
DB	916	ACGCCAGAGGTAGCTGTCCTCTCATGTGAACCATTAACCACTTATCCCTTATGAACAA	975	
QY	320	MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339	



Db 976 ATGTCTGAATTCGAAATGGAACCAAGTTCCTCGTATTATTCCTTCCTGTTATCGTTCAACCAT 1035  
 Qy 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1036 TGGGTACAGATTCAAGACCAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095  
 Qy 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379  
 Db 1096 AGTCGACACAGCTCCAGCAATCCAAATTCAT-----GGGAAATTCGTCAAGAGACT 1149  
 Qy 380 ValArgLysValGlyGluGlyTyrValPheGluGluGlyLysIleSerArgTyrValPhe 399  
 Db 1150 GTTCGAAAGTAGCGGTATGCTCTTTGAGGAGATGAGGTTCCTCGTTATATCCCA 1209  
 Qy 400 AlalysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
 Db 1210 GCCAAGGATCTTCAGCAGAACACAGCAGCGCATTTGATAGCAAACTGGCCAAGCAGAA 1269  
 Qy 420 SerValSerHisThrLeuThrAlalysLysGluAsnValAlaProArgAspGlnGluPhe 439  
 Db 1270 AGTTTATCTATAAGCTAGCACTAAGAACTAGCACTCCCATCTAGTGTGATCGAGAAATT 1329  
 Qy 440 TyrAspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 Db 1330 TACAATAAGGCTATGACTTACTAGCAAGAATCCACAGATTTACTTGATATAAGGT 1389  
 Qy 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
 Db 1390 GCACAGTGTGATTTGAGGCTTTGGATAACCTGTGGACGACTCAAGGATGCTCAAGT 1449  
 Qy 480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 Db 1450 GATAAAGTCAAGTGTAGTGAAGATATCTTGGCTTCTAGCTCCGATTCGTATCCAGAA 1509  
 Qy 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519  
 Db 1510 CGTTTAGGAAACCAAAATGCGCAATTTACTACACTGATGATGAGATTCAGTAGCAAG 1569  
 Qy 520 LeuAlaAspLysTyrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539  
 Db 1570 TTGGCAGGCAAGTACACAGCAGAGACGGTTATATCTTTGATCCTCGTGATATAACAGT 1629  
 Qy 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAsp 559  
 Db 1630 GATGAGGGGATGCCTATGTAACTCCACATATGACCCATAGCCACTGGATTAAAAAGAT 1689  
 Qy 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579  
 Db 1690 AGTTTGTCTGAAGCTGAGAGCGGCGCCAGCTTATGCTTAAAGAGAGGTTTTCACC 1749  
 Qy 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599  
 Db 1750 CCTCTTCGACAGACCATCAGGATTCAGGAATACTGAGGCAAAAGAGCAGAGCTATC 1809  
 Qy 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619  
 Db 1810 TACAACCGMGTAAGACGAGCTTAAGAGGTGCTTATGCTATGCTTCAATCTTCAATCTTCAA 1869  
 Qy 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639  
 Db 1870 TATACCTAGAGTCAAAACCGTAGTTTATATCATCTATATGACCATTAACATAAC 1929  
 Qy 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659  
 Db 1930 ATCAAAATTTGAGTGTGTTGACGAAGGCTTTATGAGGCACCTAAGGGGTATCTCTTGAG 1989  
 Qy 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679  
 Db 1990 GATCTTTGGCGACTGTCAAGTACTATGCGAAATCCAAACGACGATCCCATTCAGAT 2049  
 Qy 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699

Db 2050 AATGTTTTGTTAGCTAGCGACCATGTTCAAAGAAACAAAATGTCAGCTGATACC 2109  
 Qy 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713  
 Db 2110 AATCAAAACGGAACCAACCAAGCGAGAGAACCTCAGACAGAAAACCTGAGGAAGAAC 2169  
 Qy 713 ----- 713  
 Db 2170 CCTCAGAGAGAACCGCAAAAGCGAGAACCCAGAGTCTCCAAAACCAACAGAGAACCA 2229  
 Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730  
 Db 2230 GAAGAATCACCAGAGGAATCAGAAACCTCAGGTTCGAGACTGAAAGCTTGACAGAAA 2289  
 Qy 731 LeuLysGluAlaGluValLeuAlaLysValThrAspSerSerLeuLysAlaAla 750  
 Db 2290 CTGAGAGAGGCTGAAGATTACTTTGGAAAAATCCAGGATCCAAATTTATCAAGTCCAATGCC 2349  
 Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnLeuThrLeuGlnIleMetAspAsnSer 770  
 Db 2350 AAAGAGACTCTCAGAGGATTAATAAATAAATTTACTATTGGCACCAGGCAACAATACT 2409  
 Qy 771 IleMetAlaGluAlaGluLysLeuAlaLeuLysGlySer 785  
 Db 2410 ATTATGGCAGAGCTGAAAAAACTATTGGCTTTATTAAAGGAGT 2454

RESULT 10  
 US-10-412-850-11  
 ; Sequence 1i, Application US/10412850  
 ; Publication NO. US20040001836A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-686  
 ; CURRENT APPLICATION NUMBER: US/10/412,850  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 11  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-850-11

Alignment Scores: 4,01e-243 Length: 2531  
 Score: 2786.50 Matches: 544  
 Percent Similarity: 77.06% Conservatives: 84  
 Best Local Similarity: 66.75% Mismatches: 140  
 Query Match: 66.90% Indels: 47  
 DB: 16 Gaps: 7

US-09-765-271-56 (1-796) x US-10-412-850-11 (1-2531)  
 Qy 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnArgValSer 19  
 Db 61 TCCTATGAGCTTGGACGCTTACCAAGCTGTGAGTAAAGAGAGTCTAATCGAGTGTCT 120  
 Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 Db 121 TATATAGTGTGATCAGGCTGTGCAAAAGCAGAAAACCTTGACACCATGAAGTCAAGT 180  
 Qy 40 LysArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrVal 59  
 Db 181 AAGAGGGAGGGGATCAACGCGCAACAAATTTGTTATCAAGATTACGGATCAAGGTTATGTG 240

QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79  
DB 241 ACCTCTCATGGAGACCATTAATCTACTATATATGCAAGGTTCCCTTATGATGCCATCATC 300  
QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
DB 301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAAGTTCAAGGATTCAGACATTGTCAAT 360  
QY 100 GluValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
DB 361 GAAATCAAGGCTGTGTATGCTAATAGGTAAACGGTAATACTATGTTTACCTTAAGGAT 420  
QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
DB 421 GCGGCTCATGCGGATAATATTCGACAAAGAGAGATTAAACGTCAGAGCGAGAACGC 480  
QY 140 SerGlnHisArgGluGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
DB 481 AGTCATAATCAACTCA-----AGAGCAGATAATGCTGTGCTGCAGCCAGAGCC 531  
QY 160 GlnGlyArgTyrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
DB 532 CAAGGACGTTATACACGATGATGGGTATATCTTCAATGCATCTGATATCATTTAGGAC 591  
QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
DB 592 ACGGCTGATGCTTATATCGTTCCTCACGCGACCATTAACCATTAATCTTCAAGATGAG 651  
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuLeuSerGlyArgGlyAsnLeuSer 219  
DB 652 TTATCAGCTAGCAGTAGTGTGCTGCGAAGCCATTATGGAATGG----- 696  
QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
DB 697 -----AAGCAGGATCTCGTCTCTTCCAAAGTTCTAGTTAATGCA 738  
QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnThrSerAsnThrAsnSer 259  
DB 739 AATCCAGCTCAACCAAGATTGTCAGAGAACCAATCTGAGTCGACCTCCACTTATCAT 798  
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
DB 799 CAA--AATCAAGGGGAAAAATTTCAAGCCCTTTAGCTGAATGTATGCTAAACCCCTTA 855  
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
DB 856 TCAGAACCCCATGTGGAAATCTGATGGCTTATTTTCAGCCAGCGAAATCACAAGTCGA 915  
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
DB 916 ACCGCCAGAGGTGTAGTGTCCCTCATGGTAAACCATTAACCATTTATCCCTTATGAACAA 975  
QY 320 MetSerGluLeuGluGluArgIleAlaArgIlePheProLeuArgTyrArgSerAsnHis 339  
DB 976 ATGTCTGAATGGAAACCAAAATGTGCTGATTTATTTCCCTTCTGTTATCCCTTCAAAACCAT 1035  
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
DB 1036 TGGGTACCATGATTCAAGACCAAGAGAACCAAGTCCACACCGACTCCAGAACCTAGTCCA 1095  
QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379  
DB 1096 AGTCCGCAACCACTCAAGCAATCCAATTGAT-----GGGAAATGTGTCAAAGAGCT 1149  
QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399  
DB 1150 GTTCGAAAAGTAGGCGATGGTATGTCTTTGAGGAGAAATGGAGTTCTCTGTTATATCCCA 1209  
QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419  
DB 1210 GCCAAGATCTTTCAGCAGAAACAGCAGAGCGGCGATTTGATAGCAAACTGGCCAGCAGGAA 1269  
QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439

DB 1270 AGTTTATCTCATAAGCTAGGAACTAGAAACTGACCTCCCATCTAGTGCAGAAATTT 1329  
QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
DB 1330 TACAATAGGCTTATGACTTACTAGCAAGATTACCAAGATTCTTCTGATTAATAAGGT 1389  
QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
DB 1390 CGACAAAGTTGATTTTGGAGCTTTGGATAACCTGTTGGAGACGACTCAAGGATGCTCAAGT 1449  
QY 480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499  
DB 1450 GATAAAGTCAAGTTAGTGGAGATATCTTGCCTTCTTAGCTCCGATTCGTCATCCAGAA 1509  
QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGln 519  
DB 1510 CGTTTAGGAAAAACCAATGCGCAATATACCTACACTGATGATGAGATTCAAGTAGCCCAAG 1569  
QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSer 539  
DB 1570 TTGGCAGGCAAGTACACAGCAGAGACGTTATATCTTTGATCCTCGTGATATACCAAGT 1629  
QY 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAsp 559  
DB 1630 GATGAGGGGGATGCTATGTAACTCCACATATGACCCATGACCCACTGGATTAAAAAAGAT 1689  
QY 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeu 579  
DB 1690 AGTTTGTCTGAGCTGAGAGCGCAGCCGAGCTTATGCTRAAGAGAAAGGTTTGACC 1749  
QY 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIle 599  
DB 1750 CCTCCTTCGACAGACCATCAGATTAGGAAATACTGAGGCAAAAGAGCAGAGACTATC 1809  
QY 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619  
DB 1810 TACAACCGMTGAAAGCAGCTTAAGAGGTGCGCACTTGATGCTGCTTACAACTCTCAA 1869  
QY 620 HisThrValGluValLysAsnGlyAsnLeuIleIleProHisLysAspHisTyrHisAsn 639  
DB 1870 TATCTGTAGAGTCAAAAACGGTAGTTTATCATCTCATTATGATACCATTTACCAATAC 1929  
QY 640 IleLysPheAlaTrpPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659  
DB 1930 ATCAAAATTTGAGTGGTTTGACGAAGGCTTTATGAGSCACCTAAGGGGTATATCTTTGAG 1989  
QY 660 AspLeuPheAlaThrIleLysTyrTyrValGluHisProAspGluArgProHisSerAsn 679  
DB 1990 GATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACTCGCATTCAGAT 2049  
QY 680 AspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699  
DB 2050 AATGGTTTGGTAACCGTAGCGACCAATGTTCAAGAAACAAAAAATGGTCAAGCTGATACC 2109  
QY 700 AsnLysAsnPhelysAlaAspGluGlu-----ProValGluGluThr 713  
DB 2110 AATCAAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAGAAACC 2169  
QY 713 ----- 713  
DB 2170 CCTCGAAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAAAACCAACAGAGGAACCA 2229  
QY 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730  
DB 2230 GAAGAAATCACAGAGGAAATCAGAAAGAACCTCAGTCGACACTGAAAGAGTTGAAGAAAAA 2289  
QY 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750  
DB 2290 CTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGGATCCAAATATCAAGTCCCATGCC 2349  
QY 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770

Db 2350 AAGAGACTCTCAGAGATTAAATAAATAATTACTATTGGCACCAGGACCAATACT 2409  
 QY 771 ILeMetAlaGluAlaGluLysLeuLeuAlaLeuLeuLysGlySer 785  
 Db 2410 ATTATGGCAGAGCTGAATAACTATTGGCTTTATTAAGGAGAGT 2454

# RESULT 11

US-10-387-783-11  
 ; Sequence 11, Application US/10387783  
 ; Publication No. US20040005331A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-683  
 ; CURRENT APPLICATION NUMBER: US/10/387,783  
 ; CURRENT FILING DATE: 2003-03-13  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 11  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-387-783-11

Alignment Scores:  
 Pred. No.: 4,016-243 Length: 2531  
 Score: 2786.50 Matches: 544  
 Percent Similarity: 77.06% Conservative: 84  
 Best Local Similarity: 66.75% Mismatches: 140  
 Query Match: 66.90% Indels: 47  
 DB: 16 Gaps: 7

US-09-765-271-56 (1-796) x US-10-387-783-11 (1-2531)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19  
 Db 61 TCTTATGAGCTGGAGCTTACCAAGCTGGTCAGGATAAGAAAGAGTCTAATCGAGTTGCT 120  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 Db 121 TATATAGATGGTGATCAGGCTGGTCAAAAGGCGCAAAACTTGACACAGATGAAGTCAGT 180  
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 Db 181 AAGAGGGAGGGGATCAACCGCAACAAATTTGTTATCAAGATTACGATCAAGTTATGTG 240  
 QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIle 79  
 Db 241 ACCTTCATGGAGACCATATATCACTACTATAATGGCAAGTTCTTATGATGATGATCATC 300  
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 Db 301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGATTGACACATTTGTCAT 360  
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
 Db 361 GAAATCAAGGGTGGTATGTCATTAAAGTAAACCGTAAATACTATGTATTACCTTAAGGAT 420  
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
 Db 421 GCRCTCATGGGATATATTCGGACAAAAGAGAGATTAACCGTCAGAAGCAGGAACGC 480  
 QY 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 Db 481 AGTCATAATCATATACTCA-----AGACAGATAATGCTGTGTCAGCCAGAGCC 531

QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 Db 532 CAAGGAGCTTATATCAACCGATGATGGGTATATCTTCAATGCTATGATATCATTTGAGGAC 591  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 Db 592 ACGGCTGATGCTTATATCTGTTCTCAGCGGACCATTTACCATTTACATTTCTTAAGATGAG 651  
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 Db 652 TTATCAGCTAGCGAGTAGCTGCTGCAGAACCTATTGGAATGGG----- 696  
 QY 220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValPro 239  
 Db 697 -----AAGCAGGATCTCTCTCTTCAAGTTCTAGTTATATAATGCA 738  
 QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 Db 739 AATCAGCTCAACCAAGATTGTCAGAGAACACAACTCTGACTGTCACCTCAACTTATCAT 798  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 Db 799 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCTTAA 855  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 Db 856 TCAGAACGCCATGTGGAATCTGATGGCCTTATTTTCGACCCAGCCAAATCAAGTCA 915  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 Db 916 ACCGCCAGAGGTGTAGCTGCTCCTCATGTTGAACATTACCATTTATCCCTTATGAACAA 975  
 QY 320 MetSerGluLeuGluGluAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339  
 Db 976 ATGCTCTGAATGGAAACAGATTGCTGCTATTAATTCCTTCGTTATCGTTCAACCAAT 1035  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 Db 1036 TGGGTACCAGATTCAAGACCAAGAACCAAGTCCACACCCAGCTCCAGACCTAGTCA 1095  
 QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379  
 Db 1096 AGTCCGCAACCAAGCTTCCAGCAATCCAATTGAT-----GGGAAATTTGTCAAAGAGCT 1149  
 QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399  
 Db 1150 GTTCGAAAAGTAGGCGATGTTATGCTTTTGAAGAGATGGAGTTTCTCGTTATATATCCCA 1209  
 QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuLysSerLysLeuSerLysGlnGlu 419  
 Db 1210 GCCAAGGATCTTTCAGCAGAAAACAGCAGCAGGCGATTGATAGCAAACTGGCCAAAGCAGAA 1269  
 QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439  
 Db 1270 AGTTTATCTCATAGCTAGGAATAAGAACTAGAACCTCCCATCTAGTGTATGATGAGATTT 1329  
 QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGly 459  
 Db 1330 TACAATAAGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGTATAATAAGGT 1389  
 QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479  
 Db 1390 CGACAAAGTTGATTTTGGAGTTTGGATTAACCTGTGTGAACGATCAAGGATGCTCAAGT 1449  
 QY 480 AsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGlu 499  
 Db 1450 GATAAAGTCAAGTTAGTGGAGATATTTCTTTCCTTCTTAGCTCCGATTCGTCATCCAGAA 1509  
 QY 500 ArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValAlaGlnIleAlaGln 519  
 Db 1510 CGTTTAGAAAACCAAAATGGCAAAATTAACCTACATGATGATGATGATGATGATGATGATG 1569  
 QY 520 LeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleSer 539

```

Db 1570 TTGGCAGGCAAGTACACAGCAGAGCGGTATATCTTTGATCCTCGTGATATACCACT 1629
Qy 540 AspGluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTripleGlyLysAsp 559
Db 1630 GATGAGGGGATGCTCTATGTAACTCCACATATGACCCATAGCCACTGGATTAAAAAAGAT 1689
Qy 560 SerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyLysLeu 579
Db 1690 AGTTTGCTGAAGCTGAGAGCGCGCAGCCAGGCTTATGCTTAAGAGAAAGGTTTGACC 1749
Qy 580 ProProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaAlaIle 599
Db 1750 CCTCTTCGACAGACCATCAGGATTCAGAAATATCTGAGCAAAAGAGGAGCAAGCTATC 1809
Qy 600 TyrAsnArgValLysGlyGluLysArgIleProLeuValArgLeuProTyrMetValGlu 619
Db 1810 TACAAACCGGTAAGAGCAGCTTAAGAGGTCCTTATGATGATGCTTATGATGCTTATGAT 1869
Qy 620 HisThrValGluValLysAsnGlyAsnLeuIleLeuProHisLysAspHisTyrHisAsn 639
Db 1870 TATACCTGTAGAAGTCAAAACCGTAGTTTATCATCTCTTATGACCATTTACCATTAAC 1929
Qy 640 IleLysPheAlaTyrPheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGlu 659
Db 1930 ATCAAAATTTGAGTGGTTTGACGAAGGCTTTTATGAGGCACCTAAAGGGGTATACTCTTGAG 1989
Qy 660 AspLeuPheAlaTyrIleLysTyrValGluHisProAspGluArgProHisSerAsn 679
Db 1990 GATCTTTGGCACTGTCAAGTACTATGTCGAACATCCAAACGACGTCGCACATTCAGAT 2049
Qy 680 AspGlyTrpGluAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspPro 699
Db 2050 AATGGTTTGGTAACGCTAGCGACCATCTTCAAGAAACAAATATGTCAGCTGATACC 2109
Qy 700 AsnLysAsnPheLysAlaAspGluGlu-----ProValGluGluThr 713
Db 2110 AATCAACCGGAAACCAACGAGCGAGGAAACCTCAGACAGAAACCTGAGGAGAAAC 2169
Qy 713 ----- 713
Db 2170 CCTCGAAGAGAAACCGAAGCGAGAAACCGAGTCTCCAAACCAACAGAGGAAACCA 2229
Qy 714 -----ProAlaGluProGluValProGlnValGluThrGluLysValGluAlaGln 730
Db 2230 GAAGAAATCACAGAGGAATCAGAGAACCTCAGCTCGAGACTGAAAGGTTGAAGAAAA 2289
Qy 731 LeuLysGluAlaGluValLeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAla 750
Db 2290 CTGAGAGGCTGAAGATTTTACTTGGAAAAATCCAGGATCCAAATTCAGTCCAAATGCC 2349
Qy 751 ThrGluThrLeuAlaGlyLeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSer 770
Db 2350 AAAGAGACTCTCAGAGATTAAAAATATTTACTATTGGCCACCCAGGACAAACAATACT 2409
Qy 771 IleMetAlaGluAlaGluLysLeuLeuAlaLeuLysGlySer 785
Db 2410 ATTATGGCAGAGCTGAAGAACTATTGGCTTTATTAAGAGAGAGT 2454

```

RESULT 12

```

US-10-412-862-5
; Sequence 5, Application US/10412862
; Publication No. US20040052781A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; FILE REFERENCE: 469201-685
; CURRENT APPLICATION NUMBER: US/10/412,862
; PRIOR FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656

```

```

; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-412-862-5

Alignment Scores:
Pred. No.: 8,34e-242 Length: 2531
Score: 2772.00 Matches: 541
Percent Similarity: 75.78% Conservativeness: 91
Best Local Similarity: 64.87% Mismatches: 142
Query Match: 66.55% Indels: 60
DB: 13 Gaps: 8

```

US-09-765-271-56 (1-796) x US-10-412-862-5 (1-2531)

```

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
Db 61 TCCTATGAAGTCTGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 120
Qy 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db 121 TATATAGATGTCATCAGGCTGGTCAAAAGCGCAGAAAACTTGACACAGATGAAGTCAGT 180
Qy 40 LysArgGluGlyLysAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db 181 AAGAGGAGGAGGATCAACCGCGCAACAAATCGTCATCAAGATTACGATCAAGGTTATGTG 240
Qy 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db 241 ACCTCTCATGAGACCATTTATCTACTATTAAGCAAGGTCCTTATGATGCCATCATC 300
Qy 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsn 99
Db 301 AGTGAAGAGCTCTCATGAAGATCCGAATTATCAGTTGAAGGATTTCAGACATTTGTCAT 360
Qy 100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119
Db 361 GAAATCAAGGTCGTTATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGCAT 420
Qy 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
Db 421 GCAGCTCATCGGATTAATTCGACAAAGAAAGAGATTAAACGTCAAGAGCAGGAACAC 480
Qy 140 SerGlnHisArgGluGlyThrProArgAsnAspGlyValAlaLeuAlaArgSer 159
Db 481 AGTCATAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCGAGCCAGGCC 534
Qy 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179
Db 535 CAAAGGACGCTATACACCGGATGATGTTATATCTTCAATGCATCTGATATCATTTAGGAC 594
Qy 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 595 ACGGTCGATGCTTATATCTTCTCAGCGCGACCATTCATTACATTCCTAAGAATGAG 654
Qy 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 655 TTATCAGCTAGCGAGTTAGTCTGTCAGAAAGCTTATGGAATGGG----- 699
Qy 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
Db 700 -----AAGCAGGAGTCTCGCTCTCTTCTTCAAGTCTTCTAGTTATAATGCA 741
Qy 240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 742 AATCCAGCTCAACCAAGATTGTCAGAACCAATCTGACTGTCTCACTCCAACTTATCAT 801

```

260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyLysLeuProLeu 279  
 802 CAA---AATCAAGGGGAAACATTTCAAGCGCTTTTGAATGTATGCTTAAACCCCTTA 858  
 280 SerClnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 859 TCAGAACGCCAATGGAATCTGATGGCTTATTTCCGACCCAGCGCAATCACAAAGTCGA 918  
 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 919 ACCGCCAGAGTGTAGCTGCTCCCTCATGCTAATCAACCTTTATCCCTTATCAACAA 978  
 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 979 ATGCTGAATTTGGAATAACGAATTCCTGATTAATTCCTCCCTTCGTTATCGTTCAACCAT 1038  
 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 1039 TGGGTACAGATTTCAAGCAGAACCAACCAAGTCCCAATCGACTCCCGAACCTTAGTCCA 1098  
 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 1099 AGTCGCAACCTGACCAACCAATCTCAACAGCTCCCAAGCAATCCCAATTCGAT-----GAG 1152  
 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
 1153 AAATGGTCAAGAGAGCTGTTTCGAAGAAGTAGGCGATGTATGTTTTCGAGAGAAATGGA 1212  
 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 1213 GTTCTCGTTATATCCAGCCAGAGATCTTTCAGCAGAAACAGCAGCAGGCAATTCGATAGC 1272  
 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 1273 AAATGGCCAGCAGAGAAAGTTTATCTCATAGCTAGGAGCTTAAGAAATCACTCCACCTCCA 1332  
 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 1333 TCTAGTGTACGAGATTTTACAAATAGGCTTATAGCTTACTAGCAAGAAATTCACCAAGAT 1392  
 454 LeuPhe\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
 1393 TTACTTGATTAATAAGGTCGACAGATGATTTTGAGCTTTTGGATTAACCTGTTGAACGA 1452  
 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493  
 1453 CTCAGGATGTCCTCAAGTGAATAAGTCAAGTGTATGATGATGATATCTTCCTCTTAGCT 1512  
 494 ProlleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513  
 1513 CCGATTGCTCATCCAGAACGTTTAGGAAACCAATGCGCAATTAACCTACACTGATGAT 1572  
 514 GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533  
 1573 GAGATTCAAGTCAAGTTGGCAGGCAAGTACACAAACAGAACGCGTTATATCTTTGAT 1632  
 534 GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553  
 1633 CCTCGTATATAACCAAGTATGAGGGGATGCTATGTAACCTCAATATGACCCATAGC 1692  
 554 HisTyrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573  
 1693 CACTGATTAATAAAGATGTTTCTGAAGCTGAGAGCGGCGAGCGGCTTATGCT 1752  
 574 LysGluLysGlyIleLeuProProSerProAlaAspValLysAlaAsnProThrGly 593  
 1753 AAAGAGAAAGGTTTGACCCCTCTTCACAGACCATCAGGATTCAGGAATACTAGGCA 1812  
 594 AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613  
 1813 AAAGGAGCAAGAGCTATCTCAACCCGCTGAAGAGCAGCTAAGAGGGTGCCTATGCT 1872  
 614 LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633

1873 ATGCCCTTACAAATCTTCAATATCTGTAGAAAGTCAAAACCGTAGTTTAAATCATACCTCAT 1932  
 634 LysAspHisTyrHisAsnIleLysPheAlaTyrPheAspAspHisThrTyrLysAlaPro 653  
 1933 TATGACCATTAACCAATCAATCAATTTAGTGGTTTACGAAGGCCCTTTATGAGGCACCT 1992  
 654 AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673  
 1993 AAGGGGTATCTCTTGGAGATCTTTTGGCGACTCTCAAGTACTATGTGCAACATCCAAAC 2052  
 674 GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692  
 2053 GAACGTCGGATTCAGATAATGTTTGGTAAACCTAGCGACCATGTTCGTAATAAATAAG 2112  
 693 -----LysAspHisSerGlu----- 697  
 2113 GTAGACCAAGACAGTAAACCTGATGAAGATAAGAAACATGATGAAGTAAGTGAGCCAACT 2172  
 697 ----- 697  
 2173 CACCTGATCTGATGAAAGAGAAATCACGCTGGTTTAAATCTTTCAGCAGATAATCTT 2232  
 698 ---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu 716  
 2233 TATAAACCAACACTGATACGGAAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2292  
 717 ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal 736  
 2293 GCTGAATTCCTCAAGTAGAGAAATCTCTTATTAACGCTAAGATAGCAGATGCGGAGGCC 2352  
 737 LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrIleuAlaGly 756  
 2353 TTGCTAGAAAAGTAAACAGATCTAGTATTAGACAAAATGCTATGAGACATTTGACTGCT 2412  
 757 LeuArgAsnAsnLeuThrIleuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776  
 2413 CTAAAGAGTAGTCTTCTTCGGAACGAAAGATAATAACACTATTTTCAGCAGAGTAGAT 2472  
 777 LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal 790  
 2473 AGTCTCTTGGCTTTGTTAAAGAAAGTCAACCGGCTCCTATA 2514

RESULT 13

US-10-412-850-5  
 ; Sequence 5, Application US/10412850  
 ; Publication No. US20040001836A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Johnson, Leslie S.  
 ; APPLICANT: Adamou, John E.  
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
 ; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural  
 ; TITLE OF INVENTION: Motifs  
 ; FILE REFERENCE: 469201-686  
 ; CURRENT APPLICATION NUMBER: US/10/412,850  
 ; CURRENT FILING DATE: 2003-04-14  
 ; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/113,048  
 ; PRIOR FILING DATE: 1998-12-21  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patent In Ver. 3.0  
 ; SEQ ID NO 5  
 ; LENGTH: 2531  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-10-412-850-5

Alignment Scores:  
 Pred. No.: 8,34e-242 Length: 2531  
 Score: 2772.00 Matches: 541  
 Percent Similarity: 75.78% Conservative: 91  
 Best Local Similarity: 64.87% Mismatches: 142

Query Match:	66.55%	Indels:	60
DB:	16	Gaps:	8
US-09-765-271-56 (1-796) x US-10-412-850-5 (1-2531)			
QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---	LysGluAsnAsnArgValSer 19
DB	61	TCCTATGAACCTTGGTTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGCTCTAATCGAGTTCT	120
QY	20	TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39	
DB	121	TATATAGATGGTATCAGCGCTGGTCAAAAGAGCAGAAACTTGCACACCATGATCAAGTCACT	180
QY	40	LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59	
DB	181	AAGAGGAGGGGATCAACCGCGCAACAATCGTCATCAAGATTACGATCAAGGTTATGTG	240
QY	60	ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79	
DB	241	ACCTCTCATGGAGACCATATATCTATATAATGCAAGGTCCCTTATGATGCCATCATC	300
QY	80	SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99	
DB	301	AGTGAAGAGCTCCTCATGAAGATCCGAAATTATCAGTTGAAGAGATTGACACATTTGTCAAT	360
QY	100	GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119	
DB	361	GAATCAAGGGTGGTATGTTATCAAGGTAGATCGAATAAATACTATGTTTACCTTAAGAT	420
QY	120	AlaAlaHisAlaAspAsnValArgThrIleGluIleAsnArgGlnLysGlnGluHis 139	
DB	421	GCAGCTCATCGGATAATAATTCGGACAAAAGAGAGATTAAACGTCAAGAGCAGGAACAC	480
QY	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159	
DB	481	AGTCATAATCACGGGGTGGTTCT-----AACGATCAAGCAGTAGTTGCAGCCAGAGCC	534
QY	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179	
DB	535	CAAGGAGCGCTATACACCGATGATGGTTATCTTCAATGCATCTGATATCATTTGAGGAC	594
QY	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199	
DB	595	ACGGGTGATGCTTATATCGTTCCTCAACGGGACCATTACCATTACATTCCTAAGATGAG	654
QY	200	LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219	
DB	655	TTATCAGCTAGCCAGTTAGTCTGCTGAGAGAGCCTATTGGAATGGG-----	699
QY	220	AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239	
DB	700	-----AAGCAGGGATCTCGTCTCTCTCAAGTTCTAGTTATAATGCA	741
QY	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259	
DB	742	AATCCAGCTCAACCAAGATTGTGACAGAACCAACAATCTGACTGTCTCACTCAACTTATCAT	801
QY	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeu 279	
DB	802	CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCCCTTA	858
QY	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299	
DB	859	TCAGAACCGCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAATCAAGATCGA	918
QY	300	ThrAlaArgGlyValAlaValProHisTyrHisGlyAspHisTyrHisPheIleProTyrSerGln 319	
DB	919	ACCGCAGAGGTGATGCTGCTCCCTCATGTGTAACCATTACCCTTTATCCCTTATGACAA	978
QY	320	MetSerGluLeuGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339	
DB	979	ATGTCGTGAATTCGAAACAGAAATGCTGATATATTCCCTCTCGTTATGTTCAAAACCAT	1038
QY	340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359	
DB	1039	TGGGTACCAGATTCAAGACCAAGAACCAAGTCACAAATCGACTCCGGAACCTAGTCCA	1098
QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373	
DB	1099	AGTCGCAACCTGCACCAAAATCCTCAACAGCTCCAAAGCAATCCAAATTTGAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393	
DB	1153	AAATTGGTCAAGAGAGCTGTTGCAAAAGTAGCGATGGTTATGTTCTTTGAGAGAAATCGA	1212
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413	
DB	1213	GTTTCTCTGTTATATCCAGCCCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCTTATGATAGC	1272
QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433	
DB	1273	AAACTGCCCAAGCAGGAAAGTTTATCTCAATAGCTTAGGAGCTAAGAAACTGACCTCCCA	1332
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453	
DB	1333	TCTAGTGCATCGAGAAATTTTACATAAGCTTATGACTTACTAGCAAGAAATTCACACAGAT	1392
QY	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473	
DB	1393	TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAACGA	1452
QY	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493	
DB	1453	CTCAAGGATGTCCTCAAGTGAATAAGTCAAGTGTAGTGGATGATATTTCTTGGCTCTTATAGCT	1512
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp 513	
DB	1513	CCGATTCGTATCCAGAACGTTTAGGAAACCAAAATGCGCAAAATACCTACACTGATGAT	1572
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp 533	
DB	1573	GAGATTCAGTAGTCCCAAGTTGGCAGGCAAGTACACAAACAGACGCTTATATCTTTGAT	1632
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer 553	
DB	1633	CCTCGTATATACACAGTATGAGGGGATGCTCTATGTAACTCCACATATGACCCATAGC	1692
QY	554	HisTyrIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr 573	
DB	1693	CACGTGATTAAAAAAGATAGTTGTTCTGAAGCTGAGAGCGGCGCCAGGCTTATGCT	1752
QY	574	LysGlyLysGlyIleLeuProSerProAspAlaAspValLysAlaAsnProThrGly 593	
DB	1753	AAAGAAAGAGTTTGACCCCTCTTCACAGACCATCAGGATTCAGGAAATCTAGAGGCA	1812
QY	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg 613	
DB	1813	AAAGAGCAGAGACTATCTACACCGCGTGAAGCAGCTAAGAGAGTCCCATCTGATCGT	1872
QY	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis 633	
DB	1873	ATGCTTACAACTCTCAATATATCTAGAGTCAAAACCGTAGTTTATATCATCTCAT	1932
QY	634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro 653	
DB	1933	TATGACCATTAACATAACATCAAAATTTGAGTGGTTTGCAAGAGGCTTTTATGAGCACCT	1992
QY	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp 673	
DB	1993	AAAGGGTATACCTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTCGAAACATCAAA	2052
QY	674	GluArgProHisSerAsnAspGlyTyrGlyAsnAlaSerGluHisValLeuGlyLys--- 692	
DB	2053	GAACGTCGCAATTCAGATAATGGTTTGGTAAACGCTAGCAGCATGTTGCTGTAATAAATAAG	2112
QY	693	-----LysAspHisSerGlu-----	697



60 ThrSerHisGlyAspHisTyrHisTyrTrpAsnGlyLysValProTyrAspAlaIlelle 79  
241 ACCTCTCATGGAGACCAATATCATTTACTATATATGCAAGGTCCCTTATGATGCCATCATC 300

80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
301 AGTGAAGAGCTCCTCATGAAGATCCGAATATATCAGTTGAAGGATTCAGACATTTGTTCAAT 360

100 GluValLysGlyGlyTyrValLleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
361 GAAATCAAGGGTGGTTATGTTATCATCAAGGTAGATGGAATACTATGTTTACCTTAAAGTAT 420

120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139  
421 GCAGCTCATGCCGATATATTCGRCAAAAGAAGAGATTAAACGTCAGACGAGGAACAC 480

140 SerGlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
481 AGTCATAAATCACGGGGGTGGTTCT-----AACGATCAAGCAGTATGTTGACCCAGAGCC 534

160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
535 CAAGGACGCTATACAACGGATGATGTTATATCTTCAATGCATCTGATATCATTTAGGAC 594

180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
595 ACGGTGATGCTTATATCGTTCTCTACGGCGACCATACCATACATTCCTAAGAATGAG 654

200 LeuSerAlaSerGluLeuAlaIleAlaGluAlaPheLeuSerClyArgGlyAsnLeuSer 219  
655 TTATCAGCTAGCAGTTAGTCTGTCAGAGCCCTATTGGAATGG----- 699

220 AsnSerArgThrTyrArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
700 -----AAGCAGGGATCTCGTCTCTTCAAGTTCATGTTATATATGCA 741

240 SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
742 AATCCAGCTCAACCAGATTTGTCAGAGAACCAACAATCTGACTGCTCACTCCAATTTATCAT 801

260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
802 CAA---AATCAAGGGGAAAACATTTCAAGCCTTTTACGTGAATGTATGCTAAACCTTTA 858

280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
859 TCAGAACGCCATGTGGAATCTGATGGCTTATTTTCGACCCAGCGCAAAATCACAAGTCGA 918

300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
919 ACCGCGCAGAGTGTAGCTGTCCCTCATGTGAACCAATACCACTTTATCCCTTTATGAACAA 978

320 MetSerGluLeuGluAlaArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
979 AGTCTGHAATTTGGHAAAACGAATTCCTCGTATTTATTCCTCTGTTATCTTCAACCAT 1038

340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
1039 TGGGTACCAGATTCAGACACAGAACCAACAGTCCCAATCGACTCCGGAACCTTAGTCCA 1098

360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
1099 AGTCGCGCAACCTTGCACCAAAATCTCTCAACGACTCAAGCAATCCAATGAT-----GAG 1152

374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
1153 AAATTTGGTCAAAGAAGCTGTTTCAAAAAGTAGCGCATGTTATGTCTTTTAGGAGAAATGGA 1212

394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
1213 GTTTTCGTATATCCAGCCCAAGGATCTTTCAGCGAAGAACAGCAGCGCAATGTGATGAC 1272

RESULT 14  
US-10-387-783-5  
; Sequence 5, Application US/10387783  
; Publication No. US20040005331A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Leslie S.  
; APPLICANT: Adamou, John E.  
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus  
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural

200	LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheIeuSerGlyArgGlyAsnLeuSer	219
Db	TTATCAGCTACGAGTGTAGCTGTCAGAGAAGCCTATTGGAATGGG-----	699
220	AsnSerArgThrTyArgAGGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro	239
700	-----AAGCAGGGATCTCGTCCCTCTTCAAGTCTAGTTATAATGCA	741
240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259
742	AATCCAGCTCAACCAAGATTGTCAGAGAACCAACAATCTGACTGCTCACTCCAACCTTATCAT	801
260	GlnAlaSerGlnSerAsnAspIleAaspSerLeuLeuLeuGlnLeuTyryLysLeuProLeu	279
802	CAA---AATCAAGGGGAAAAACATTCAAGGCTTTTACGTAAGTATGATGCTAAACCCCTTA	858
280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299
859	TCAGAACGCCATGTGGAACTGTATGGGCTTATTTCGACCCAGCGCAATATCACAAGTCGA	918
300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319
919	ACCGCCAGAGGTGTAGTGTGCTCCCTCATGTGAACCAATTACCACCTTATCCCTCTATGACCAA	978
320	MetSerGluLeuGluGluAargIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339
979	ATGTCCTGAATTGGAAAAACGAATGCTCGTATTATTCCTCTCGTTATCGTTCATCGTTCAAACCAT	1038
340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro	359
1039	TGGGTACCAGATTCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTTAGTCCA	1098
360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
1099	AGTCGCGCAACCTGCACCAAAATCCTCAACCAAGCTCAAGCAATCCAAITGAT-----GAG	1152
374	SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
1153	AAATTGGTCAAGAAGCTGTTCCGAAAGTAGCGATGGTTATGCTTTGAGGACAATGGA	1212
394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
1213	GTITTCGTATTATCCCAAGCATCTTTTCAGCAGAAACAGCAGCAGGAGGATGATGAGC	1272

QY	320	MetSerGluLeuGluGluArgIlleAlaArgIlleIleProLeuArgTyrArgSerAsnHis	339
DB	979	ATGTCGAATTGGAAAAACGAATTGCTCGTATTATTCCTTGGTATCGTTCAAAACCAT	1038
QY	340	TrpValProAspSerArgProGluGluProSerProGlnProThrProGluProSerPro	359
DB	1039	TGGGTACCAGATTCAAGACCAGACACCAAGTCCCAATCGACTCCGGAACCTAGTCCA	1098
QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
DB	1099	AGTCGCGCAACCTGCACCAAAATCTCAACGAGCTCCAAGCAATCAAAATGTAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValAlaArgLysValGlyGluGlyTyrValPheGluGluLysGly	393
DB	1153	AAATTGGTCAAAAGAGCTGTTCCGAAAGTAGCGATGTTATGCTTTGAGGACAATGGA	1212
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
DB	1213	GTTTCTCGTTATATCCAGCCAAAGGATCTTTTCAGCAGAAAACAGCAGCAGGCATTGATAGC	1272

QY	360	GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer	373
		:::	
Db	1099	AGTCGCGAACCTGCACCAATCCTCAACGAGCTCCAAGCAATGAT-----GAG	1152
QY	374	SerLeuValSerGlnLeuValIargLysValGlyGluGlyTyrValPheGluGluLysGly	393
		:::	
Db	1153	AAATTGGTCAAGAAGCTGTTCCGAAAGTAGCGCATGTTATGTTCTTTGAGGAGAATGGA	1212
QY	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413
		:::	
Db	1213	GTTTCTCGTTATATCCGACCAAGGATCTTTTCAGCAGAAACAGCAGCAGGCGATTTGATAGC	1272

Db	1099	AGTCCGCAACCTGCACCAAACTCTCAACGAGCTCAAGCAATCTCAATTGAT-----GAG	1152
Qy	374	SerLeuValSerGlnLeuValArgIysValGlyGluGlyTyrValPheGluGluLysGly	393
Db	1153	AAATTGGTCAAGAAGCTGTTCGAAAGTAGCGCATGGTTATGCTTTGAGGAGAAATGGA	1212
Qy	394	IleSerArgTyrValPheAlaIysAspLeuProSerGluThrValLysAsnLeuLysSer	413
Db	1213	GTTTCGTTATATCCGACCAAGGATCTTTAGCAGAAACAGCAGCGCGGCGATTGATAGC	1272



QY	414	LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLeuLysLysGluAsnValAla	433
DB	1273	AAACTGGCCAGCAGCAAGATTATCTCATAAGCTAGGAGCTAAGAAACCTGACCTCCCA	1332
QY	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLeuAla	453
DB	1333	TCTAGTGATCGAGAATTTTCAANTAAGCGTTATGACTTACTTAGCAAGAAATTCACCAAGAT	1392
QY	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473
DB	1393	TTACTTTGATAATAAAGGTCGCACAGTTGATTTTGAGGCTTTGGATAACCTGTTGGAAACGA	1452
QY	474	LeuAsnAspGluSerThrAsnLysGlnLysLeuValAspLeuLeuAlaPheLeuAla	493
DB	1453	CTCAAGGATGTCCTCCCAAGTCATAAAGTCAAGTTAGTGGATGATATCTTGCTCTTAGCT	1512
QY	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
DB	1513	CGATTTCGTCTCAGAACGTTTAGGAAACCAATGCGAAATTCCTACACTGATGAT	1572
QY	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
DB	1573	GAGATTCAAGTAGCCAAGTTGGCAGGCAGTACACAACAGACGCGTTATATCTTTGAT	1632
QY	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
DB	1633	CCTCGTGATATAACACAGTGATGAGGGGATGCGCTTATGAACTCCACATATGACCCATAGC	1692
QY	554	HisTrpIleGlyLysAspSerLeuSerAspLysGlnLysValAlaGlnAlaTyrThr	573
DB	1693	CACTGGATTAAAAAGATAGTTTCTCGAAGCTGAGAGCGCAGCCAGCGCTTATGCT	1752
QY	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
DB	1753	AAAGAGAAAGTTTGACCCCTCTTCACAGACCATCAGGATTCAGAAATACTGAGGCA	1812
QY	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyGluLysArgIleProLeuValArg	613
DB	1813	AAAGGACGAGAAGCTATCTACACCGCGTGAAGCAGCTAAGAAGGTGCCACTTGATCGT	1872
QY	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIlelleProHis	633
DB	1873	ATGCCTTTACAATCTTCAATATATCTGTAGAAGTCAAAACCGTAGTTTAAATCATCCTCAT	1932
QY	634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro	653
DB	1933	TATGACCATTTACCATAACATCAAAATTTGAGTGGTTTGACGAGGCCCTTTATGAGGCACCT	1992
QY	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp	673
DB	1993	AAGGGGTATACTCTTGAGGATCTTTTGGGACTGTCAAGTACTATGTGCAACATCCAAAC	2052
QY	674	GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys---	692
DB	2053	GAACGTCGCATTTCAGATATGTTTTTGTAAACGTAGCGACATGTTCTGTAATAATAAG	2112
QY	693	-----LysAspHisSerGlu-----	697
DB	2113	GTAGACCAACAGACAGTAAACCTGTATGAAGATAGGAACATCATGAAGTAGTGACCAACT	2172
QY	697	-----	697
DB	2173	CACCCCTGAATCTGATGAAAAAGAGAAATCACGCTGGTTTTAAATCTTCACGAGATAATCTT	2232
QY	698	---AspProAsnLysAsnPheLysAlaAspGluGluProValGluGluThrProAlaGlu	716
DB	2233	TATAAACCAAGCACTGATACGGAAGACAGAGGAAGAGCTGGAAGATACCAAGATGAG	2292
QY	717	ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal	736
DB	2293	GCTGAAATTCCTCAAGTAGAGAATCTCTGTATTAAACGTAGATAGACAGATCGCGAGGCC	2352
QY	737	LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly	756

```

Db      2353 TTGCTAGAAAAAGTAACAGATCCCTAGTATTAGACAAAATGCTATGGAGACATTGCACTGGT 2412
QY      757  LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu 776
Db      2413 CTAAAAAGTAGTCTTCTCTCGGACGAAGATATAACACTATTCAGCAGAGATGAT 2472
QY      777  LysLeuLeuAlaLeuLeuLysGlySerAsnProSerVal 790
Db      2473 AGTCTCTGGCTTTGTTAAAAAGAAAGTCAACCGGCTCCTATA 2514

RESULT 15
US-09-884-465A-5
; Sequence 5, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Alignment Scores:
Pred. No.: 7,96e-240 Length: 2639
Score: 2750.50 Matches: 538
Percent Similarity: 75.30% Conservative: 90
Best Local Similarity: 64.51% Mismatches: 145
Query Match: 66.04% Indels: 61
DB: 10 Gaps: 8

US-09-765-271-56 (1-796) x US-09-884-465A-5 (1-2639)
QY      1  SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19
Db      174 TCTATGAACCTGGTCGTCCACCAAGCTGGTCAGGTTAAGAAAGATCTAATCGAGTTTCT 233
QY      20  TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
Db      234 TATATAGATGGTGATCAGAGCTGGTCAAAAGCGCAGAAATTTGACACCAAGATGAAGTCAGT 293
QY      40  LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
Db      294 AAGAGAGAGGGGATCAACGCCGAAACAAATTGTTATCAAGATTACGGATCAAGTTATGTG 353
QY      60  ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIle 79
Db      354 ACCTCTCATGGAGACCATTATCATTTACTATATAATGGCAAGGTTCTTTATGTCATCATC 413
QY      80  SerGluGluLeuLeuMetIleAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99
Db      414 AGTGAAGAACTTCTCATGAAGATCCGAATTATCAAGTTGAAGGATTCAGACATTGTCAAT 473
QY      100 GluValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLysLeuAsp 119
Db      474 GAAATCAAGGGTGGCTATGTGATTAAAGGTACACGGAATAACTACTATGTTTACCTTAAAGAT 533
QY      120 AlaAlaHisAlaAspAsnValArgThrLysGluGlnIleAsnArgGlnLysGlnGluHis 139
Db      534 GCGGCCCATCGGCAATATTTCGCAACAAGAGAGATTAAACGTCAAGACGACGAACAC 593

```

Qy	140	SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer	159	Qy	494	ProIleThrHisProGluArgLeuGlyLysProAsnSerGlnIleGluTyrThrGluAsp	513
Db	594	AGTCAATATCAATACTCA-----AGACGATATAATGCTGTGTCTGCAGCAGAGCC	644	Db	1623	CCGATTCGTTCATCCAGAACGTTTAGGAAACCAATCGCAATACCTACCTACGATGAT	1682
Qy	160	GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleGluAsp	179	Qy	514	GluValArgIleAlaGlnLeuAlaAspLysTyrThrThrSerAspGlyTyrIlePheAsp	533
Db	645	CAAGGACGTTATACAACGGATGATGGGTATATCTTCATGTCATCTGATATCAATTGAGGAC	704	Db	1683	GAGATTCAAGTAGCAAGTTGGCAGGAGTACACACAGAGAGCGGTATATCTTTGAT	1742
Qy	180	ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProGlyAsnGlu	199	Qy	534	GluHisAspIleIleSerAspGluGlyAspAlaTyrValThrProHisMetGlyHisSer	553
Db	705	ACGGGTGATCTTATATCTTCTCACGGCACCATTACCATTAATCTCTTAAGAATGAG	764	Db	1743	CCTCGTATATAACCATGATGAGGGGATCGCTTATGTAACTCCACATATGACCATAGC	1802
Qy	200	LeuSerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer	219	Qy	554	HisTrpIleGlyLysAspSerLeuSerAspLysGluLysValAlaAlaGlnAlaTyrThr	573
Db	765	TTATCAGTAGCAGTTAGCTGCTGCAGAACCTTATGGAAATGG-----	809	Db	1803	CACCTGGATTAAAAAAGATAGTTTCTCTCAACTGAGAGCGGCGCCGCTTATGCT	1862
Qy	220	AsnSerArgThrTyrArgGluAsnSerAspAsnThrSerArgThrAsnTrpValPro	239	Qy	574	LysGluLysGlyIleLeuProProSerProAspAlaAspValLysAlaAsnProThrGly	593
Db	810	-----AAGCAGGGATCTCGTCTCTTCTCAAGTCTTAGTTATAATGCA	851	Db	1863	AAAGAGAAAGTTTGAACCCCTCTTCGACACACACAGGATTGAGGAAATACTGAGGCA	1922
Qy	240	SerValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSer	259	Qy	594	AspSerAlaAlaAlaIleTyrAsnArgValLysGlyLysArgIleProLeuValArg	613
Db	852	AATCCAGTTCAACCAAGATTGTACAGAAACCAACATCTCACTGTCTCACTCCCACTTATCAT	911	Db	1923	AAAGAGCAGAGAGCTATCTACAAACCGCTGAAAGCAGCTAAAGAGGTGCCACTTGTATCGT	1982
Qy	260	GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu	279	Qy	614	LeuProTyrMetValGluHisThrValGluValLysAsnGlyAsnLeuIleIleProHis	633
Db	912	CAA---AATCAAGGGGAAACATTTCAAGCCCTTTACGTGAATGTATGTCTAAACCCCTTA	968	Db	1983	ATGCTTCAATCTTCAATATATCTAGAGAGTCAAAACCGGTAGTTTATCTACTCTCAT	2042
Qy	280	SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg	299	Qy	634	LysAspHisTyrHisAsnIleLysPheAlaTrpPheAspAspHisThrTyrLysAlaPro	653
Db	969	TCAGAACGCCATGTAGATCTGATGCCCTTATTTTCGACCCAGCGCAATCAACAGTCGA	1028	Db	2043	TATGACCATTTACCATTAACATCAATTTGAGTGTGTTGACGAGGCGCTTTATGAGCACCT	2102
Qy	300	ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln	319	Qy	654	AsnGlyTyrThrLeuGluAspLeuPheAlaThrIleLysTyrTyrValGluHisProAsp	673
Db	1029	ACCGCCAGAGGTAGTCTGCTCATGTGTAAACCATTAACCTTTATCCCTTATGAACAA	1088	Db	2103	AAAGGGTATAGTCTTGGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAC	2162
Qy	320	MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis	339	Qy	674	GluArgProHisSerAsnAspGlyTrpGlyAsnAlaSerGluHisValLeuGlyLys---	692
Db	1089	ATGTCGTAATGGAAAGAAAGAAATGCTGTATATTTCCCTTCTGTTATGTTCAACCAT	1148	Db	2163	GAACGTCGCAATTCAGATAATGTTTGTGTAACGCTAGTGACCATGTTGCTAAATAAG	2222
Qy	340	TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSer---	358	Qy	693	-----LysAspHisSerGlu-----	697
Db	1149	TGGGTACCATGATCAAGACCAAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA	1208	Db	2223	GCAGACCAAGATAGTAAACCTGATGAAGATAAGGAACATGATGAAGTAAGTGAGCCAACT	2282
Qy	359	-----ProGlyProGlnProAlaProAsnLeuLysIleAspSerAsnSer	373	Qy	697	-----	697
Db	1209	AGTCTGCAACCTGCACCAAAATCTCAACAGCTCCCAAGCAATCCAATTGAT-----GAG	1262	Db	2283	CACCTGTAATCTGATGAAAAAGAGAATCACGCTGTTTAAATCTTTCAGCAGATAATCTT	2342
Qy	374	SerLeuValSerGlnLeuValArgLysValGlyGlyTyrValPheGluGluLysGly	393	Qy	698	---AspProAsnLysAsnPhelysAlaAspGluGluProValGluGluThrProAlaGlu	716
Db	1263	AAATGGTCAAGAACTGTTTCAAAAGTAGGCGATGTTATGTCTTTGAGGAGATGGA	1322	Db	2343	TATAAACCAAGCACTGATACGGAAGAGACAGAGGAAGAGAGCTGAAGTATCCACAGATGAG	2402
Qy	394	IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer	413	Qy	717	ProGluValProGlnValGluThrGluLysValGluAlaGlnLeuLysGluAlaGluVal	736
Db	1323	GTTCCTCGTTATATCCAGCCAGGATCTTTCAGCAGAAACACAGCAGGAGCATTTGATAGC	1382	Db	2403	GCTGAAATTTCTCAAGTAGAGAAATCTGTTTAACTGCTAAAGTAGCAGATGCGGAGGCC	2462
Qy	414	LysLeuSerLysGlnGlnSerValSerHisThrLeuThrAlaLysLysGluAsnValAla	433	Qy	737	LeuLeuAlaLysValThrAspSerSerLeuLysAlaAsnAlaThrGluThrLeuAlaGly	756
Db	1383	AAACTGCCAAGCAGGAAAGTTTATCTCAAACTAGGAGCTAAGAAACCTGACCTCCCA	1442	Db	2463	TTGCTAGAAAAAGTAACAGATCTCTAGTATTAGACAAAATGCTATGGAGACATTGACTGGT	2522
Qy	434	ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla	453	Qy	757	LeuArgAsnAsnLeuThrLeuGlnIleMetAspAsnAsnSerIleMetAlaGluAlaGlu	776
Db	1443	TCATGATGATCGAATTTTACATTAAGCTTATGACTTACTAGCAAGATTCACCAAGAT	1502	Db	2523	CTAAAAAGTAGTCTTCTTCTCGGAACGAAAGATAATAACACTATTTTCAGCAGAGTAGAT	2582
Qy	454	LeuPhe***AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg	473	Qy	777	LysLeuLeuAlaLeuLeuLysGlySerAsnProSerSerVal	790
Db	1503	TTACTTGATATAAAGTTCACAAAGTTGATTTTGGAGGTTTGGATAACCTGTTGGACGA	1562	Db	2583	AGTCTCTGGCTTTGTTAAAAAGAAAGTCAACCGGCTCCTATA	2624
Qy	474	LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla	493	Db			
Db	1563	CTCAAGGATGTCTCAAGTGATAAAGTCAAGTTAGTGTGATGATATTTCTGCTTCTTAGCT	1622	Db			

RESULT 16  
US-09-765-272-65  
; Sequence 65, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:

; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/765,272  
 ; FILING DATE: 22-Jan-2001  
 ; CLASSIFICATION: <Unknown>  
 ;  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/961,083  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ;  
 ; INFORMATION FOR SEQ ID NO: 65:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2290 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ;  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 65:

# Alignment Scores:

Pred. No.: 9.94e-231 Length: 2290  
 Score: 2649.50 Matches: 519  
 Percent Similarity: 76.51% Conservative: 77  
 Best Local Similarity: 66.62% Mismatches: 130  
 Query Match: 63.61% Indels: 53  
 DB: 9 Gaps: 8

US-09-765-271-56 (1-796) x US-09-765-272-65 (1-2290)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrVal---LysGluAsnAsnArgValSer 19  
 DB 5 TCCTATGAACCTTGCTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCT 64  
 QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39  
 DB 65 TATATAGATGGTGATCAGGCTGTGTCAAAAGCGCAGAAACTTGACCCAGATGAAGTCAGT 124  
 QY 40 LysArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59  
 DB 125 AAGAGGGAGGGGATCAACCGCGACCAATNGTNATCAAGATTACGGATCAAGGTTATGTG 184  
 QY 60 ThrSerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIle 79  
 DB 185 ACCTCTCATGGACCATATATCATTAATAATGGCAAGGTTCTTATGATGCATCAIC 244  
 QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsn 99  
 DB 245 AGTGAAGAGCTCTCATGAAGAAATCGGAATTAATCAGTTGAAGGATTTCAGACATTGCAAT 304  
 QY 100 GluValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAsp 119  
 DB 305 GAATCAAGGGTGGTTATGTCTATAGGTAACGGTAAATATCTATGTTACCTTAAAGGAT 364  
 QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139

DB 365 GCAGCTCATGGGATAATAATTCGGACAAAAGAGATTAAACGTCAGAAAGCAGAAACGC 424  
 QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159  
 DB 425 AGTCATATATCACTCA-----AGACAGATATACTGTTGTTGCTGCAGCCAGAGCC 475  
 QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAsp 179  
 DB 476 CAAGGACGTTATACCAACGATGATGGGTATATCTTCAATGCATCTGATATCATTCAGGAC 535  
 QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199  
 DB 536 ACGGTGATGCTTATATCGTTCTCCACGCGACCATTAACCATTAATTCCTAAGAATCAG 595  
 QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219  
 DB 596 TTATCAGCTACGAGTTAGCTGCTGCAGAACCTTATGGAATGGG----- 640  
 QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239  
 DB 641 -----AAGCAGGATCTCGTCTCTTCAAGTTCTAGTTATAATGCA 682  
 QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259  
 DB 683 AATCCAGCTCAACCAAGATTGTCAGAAACCAATCTGACTGCTCACTCCAACTTATCAT 742  
 QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279  
 DB 743 CAA---AATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATTTGATGCTAAACCTTTA 799  
 QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299  
 DB 800 TCAGACGCCATGTGGAATCTGATGGCTTATTTTCGACCAGCGCAATCACAAGTCGA 859  
 QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319  
 DB 860 ACCGCCAGAGGTGATGCTGCTCCCTCATGTTAAACCAATACCACCTTATCCCTTATGAACAA 919  
 QY 320 MetSerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHis 339  
 DB 920 ATGCTGTAATGGAAACCAATTCGCTGTTATTTCCCTTCTGTTATCGTTTCAACCAT 979  
 QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359  
 DB 980 TGGGTACCAATTCAAGACCAGAACCAACCAAGTCCCAATCGACTCCGGAACCTAGTCCA 1039  
 QY 360 GlyProGlnProAlaProAsnLeuLys-----IleAspSerAsnSer 373  
 DB 1040 AGTCCGCACTCCACCAATCTCAACCAAGCTCCCAAGCAATCCAATTGAT-----GAG 1093  
 QY 374 SerLeuValSerGlnLeuValArgLysValGlyGluGlyTyrValPheGluGluLysGly 393  
 DB 1094 AAATTGGTCAAGAACCTGTTTCGAAAAGTAGGCGATGGTTATGTTCTTTGAGGAGAATGGA 1153  
 QY 394 IleSerArgTyrValPheAlaLysAspLeuProSerGluThrValLysAsnLeuGluSer 413  
 DB 1154 GTTCTCTGTTATATCCCAAGGATCTTTCAGAGAAACACGACGACGACGATTTGATAGC 1213  
 QY 414 LysLeuSerLysGlnGluSerValSerHisThrLeuThrAlaLysLysGluAsnValAla 433  
 DB 1214 AAATCGCCACGACGAGAAAGTTTATCTCAATAGCTTATGACTTACTAGCAAGATTCCACCAAGAT 1273  
 QY 434 ProArgAspGlnGluPheTyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAla 453  
 DB 1274 TCTAGTATCGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGATTCCACCAAGAT 1333  
 QY 454 LeuPhe\*\*\*AsnLysGlyArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArg 473  
 DB 1334 TTACTTGATAATAAAGGTCGACAAAGTTGATTTTGAGCTTTTGATAACCTGTTGGNACGA 1393  
 QY 474 LeuAsnAspGluSerThrAsnLysGluLysLeuValAspAspLeuLeuAlaPheLeuAla 493

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 243:
; US-10-158-844-243
;
; Alignment Scores:
; Pred. No.: 4, 02e-151 Length: 2359
; Score: 1774.00 Matches: 347
; Percent Similarity: 81.22% Conservativeness: 51
; Best Local Similarity: 70.82% Mismatches: 74
; Query Match: 42.59% Indels: 18
; DB: 13 Gaps: 5
;
; US-09-765-271-56 (1-796) x US-10-158-844-243 (1-2359)
;
; QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
; DB 940 TCCTATGAGCTTGGAGCTTACCAAGCTGGTCAGGATAGAAAGAGCTCTAATCGAGTTGCT 999
;
; QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
; DB 1000 TATATAGATGGTGCATCAGCTGGTCAAAAGGCGAGAAATCTTACACCAAGTGAAGTCAGT 1059
;
; QY 40 LysArgGluGlyLysAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
; DB 1060 AAGAGGAGGGGATCAACCGCCGACAAATTTGTTATCAAGATTACGGATCAAGGTTATGTG 1119
;
; QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIlelle 79
; DB 1120 ACCTCTCATGGAGACCATTTATCTACTATAATGGCAAGGTTCTCTTATGATGCCATCATC 1179
;
; QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsn 99
; DB 1180 AGTGAAGAGCTCTCATGAAGAGATCCGAATTTACGTTGAGGATTCAGACATTTGTCAT 1239
;
; QY 100 GluValLysGlyLysTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119
; DB 1240 GAAATCAGGGTGGTTATGTCTTAAGGTAACCGTAAATATACTATCTTTTACCTTAAGGAT 1299
;
; QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
; DB 1300 GCAGCTCATGGGATTAATATTCGGACAAAAGAGAGATTAAACGTGAGAGCAGGACGCC 1359
;
; QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
; DB 1360 AGTCATAATCATTAATCA-----AGAGCAGATAATCTGTGTGCTGAGCCAGAGGCC 1410
;
; QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIlelleGluAsp 179
; DB 1411 CAGGACGTTTATACACGGGATGATGGGTATATCTTCAATGCTCATCTGATATCATTCAGGAC 1470
;
;
; RESULT 17
; US-10-158-844-243
; Sequence 243, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 243:
; US-10-158-844-243
;
; Alignment Scores:
; Pred. No.: 4, 02e-151 Length: 2359
; Score: 1774.00 Matches: 347
; Percent Similarity: 81.22% Conservativeness: 51
; Best Local Similarity: 70.82% Mismatches: 74
; Query Match: 42.59% Indels: 18
; DB: 13 Gaps: 5
;
; US-09-765-271-56 (1-796) x US-10-158-844-243 (1-2359)
;
; QY 1 SerTyrGluLeuGlyLeuTyrGlnAla---ArgThrValLysGluAsnAsnArgValSer 19
; DB 940 TCCTATGAGCTTGGAGCTTACCAAGCTGGTCAGGATAGAAAGAGCTCTAATCGAGTTGCT 999
;
; QY 20 TyrIleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSer 39
; DB 1000 TATATAGATGGTGCATCAGCTGGTCAAAAGGCGAGAAATCTTACACCAAGTGAAGTCAGT 1059
;
; QY 40 LysArgGluGlyLysAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrVal 59
; DB 1060 AAGAGGAGGGGATCAACCGCCGACAAATTTGTTATCAAGATTACGGATCAAGGTTATGTG 1119
;
; QY 60 ThrSerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIlelle 79
; DB 1120 ACCTCTCATGGAGACCATTTATCTACTATAATGGCAAGGTTCTCTTATGATGCCATCATC 1179
;
; QY 80 SerGluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsn 99
; DB 1180 AGTGAAGAGCTCTCATGAAGAGATCCGAATTTACGTTGAGGATTCAGACATTTGTCAT 1239
;
; QY 100 GluValLysGlyLysTyrValIleLysValAspGlyLysTyrTyrValTyrLysLysAsp 119
; DB 1240 GAAATCAGGGTGGTTATGTCTTAAGGTAACCGTAAATATACTATCTTTTACCTTAAGGAT 1299
;
; QY 120 AlaAlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHis 139
; DB 1300 GCAGCTCATGGGATTAATATTCGGACAAAAGAGAGATTAAACGTGAGAGCAGGACGCC 1359
;
; QY 140 SerGlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSer 159
; DB 1360 AGTCATAATCATTAATCA-----AGAGCAGATAATCTGTGTGCTGAGCCAGAGGCC 1410
;
; QY 160 GlnGlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIlelleGluAsp 179
; DB 1411 CAGGACGTTTATACACGGGATGATGGGTATATCTTCAATGCTCATCTGATATCATTCAGGAC 1470
;
;
; RESULT 17
; US-10-158-844-243
; Sequence 243, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;

```

```

QY 180 ThrGlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGlu 199
Db 1471 ACGGGTGGTGTATATCGTCTCCTCAGCGCGACCATACCATATCATCTTAAAGATGAG 1530
QY 200 LeuSerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSer 219
Db 1531 TTATCAGCTAGGAGTGTAGCTCTCAGAAAGCTATTGGAATGGG----- 1575
QY 220 AsnSerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValPro 239
Db 1576 -----AAGCAGGGATCTCTCTCTTCAAGTTCTAGTTTAAATGCA 1617
QY 240 SerValSerAsnProGlyThrThrAsnThrSerAsnAsnSerAsnThrAsnSer 259
Db 1618 AATCCAGCTCAACCAAGATGTCAGAGAACCAATCTGACTGTCTCACTCCCACTTATCAT 1677
QY 260 GlnAlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeu 279
Db 1678 CAA---AATCAAGGGGAAACATTTCAAGCCCTTTTACGTGAATGTATGCTAAACCCCTTA 1734
QY 280 SerGlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArg 299
Db 1735 TCAGAACGCCATGTGGAATCTGATGCGCTTATTTTCGACCCAGCCCAATCACAGTCCGA 1794
QY 300 ThrAlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGln 319
Db 1795 ACCGCCAGAGGTAGTGTCCCTCATGGTAACCATTAACCACTTATCCCTTATCAACAA 1854
QY 320 MetSerGluLeuGluGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHis 339
Db 1855 ATGCTGAAATGGAAACAAAGATGCTCGTATTTATCCCTTCGTTATCGTTCAAAACCAT 1914
QY 340 TrpValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerPro 359
Db 1915 TGGGTACCAAGTTCAGACCAAGAACCAAGTCCACCAACCGACTCCAGAACCTTAGTCCA 1974
QY 360 GlyProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeu 379
Db 1975 AGTCGCAACAGCTCCCAAGCAATCCAAATTGAT-----GAGAAATTTGGTCAAGAGAGCT 2028
QY 380 ValArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPhe 399
Db 2029 GTTCGAAAGTAGGCGATGTGTATGCTTTGAGGAGATGGAGTTCTCGTTATATCCCA 2088
QY 400 AlaLysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGlu 419
Db 2089 GCCAAGATCTTCAGCAGAAACAGCAGCAGGCAATTGATAGCAAACTGGCCAGCAGGAA 2148
QY 420 SerValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPhe 439
Db 2149 AGTTTATCTCATAGCTAGGAACTAAGAAACTGACCTCCCATCTAGTGTATCGAATTT 2208
QY 440 TyrAspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGly 459
Db 2209 TACAATAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTCGATAATAAGGT 2268
QY 460 ArgAsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThr 479
Db 2269 CGACAAGTTGATTTTCAGGCTTTGGATTAACCTGTGGAACGACTCAAGGATGCTCAAGT 2328
QY 480 AsnLysGluLysLeuValAspAspLeuLeu 489
Db 2329 GATAAAGTCAAGTTAGTGGAAAGATATCTT 2358

```

RESULT 18

```

US-09-884-465A-1
; Sequence 1, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis

```

```

; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3120
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-1

```

```

Alignment Scores:
Pred. No.: 4,78e-103 Length: 3120
Score: 1247.00 Matches: 298
Percent Similarity: 54.36% Conservativity: 95
Best Local Similarity: 41.22% Mismatches: 180
Query Match: 29.94% Indels: 150
DB: 10 Gaps: 19

```

```

US-09-765-271-56 (1-796) x US-09-884-465A-1 (1-3120)
QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 64 GCCTATGCCTAAACCCAGCATCGTTCGACGAGAAATAGGACATAATCGTGTCTCTAT 123
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGATGCGAGCCAGTCAGTCAAGTCAGAAAGTGAACACTTGACACAGCAGGTTAGCCAG 183
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 184 AAGAAGCAATTCAGGCTGAGCAATTTGTAATCAAAATACAGATCAGGGCTATGTAACG 243
QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 244 TCACACGGTGACCATATCATCTATAATATGAGGAAAGTTCCTTATGATGCCCTCTTAGT 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 304 GAAGAACTCTTGATGAGGATCCAAACTATCAACTTAAAGACGCTGATATTGTCATGAA 363
QY 101 ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120
Db 364 GTCAAGGTGGTTATATCATCAAGGTGATGGAAATATTTATGTCTACCTGAAAGATGCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 424 GCTCATGCTGATAATGTTGAACTAAAGATGAATCAATCGTCAAAACCAAGAACATGTC 483
QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 484 AAGATAATAG-----AAGTTAACTCTAATGTTCTGTAGCAAGGTCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 532 GGACGATATACCAAAATGATGTTAIGTCTTTAATCCAGCTGATATTATCGAAGATACG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAATGCTTATATCGTCTCATGGAGGTCACTATCATCATATCCCAAAACCGGATTTA 651
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGTAGTGAATTAGCAGCAGTAAAGCACATCTGGCTGGAAATAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATCTTCAACAGCTAGTAGCAAT----- 744

```

Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnThrAsnSerGln 260  
 Db 745 -----AACACGCAATCTCTAGCAAAAGGATCAACATAGCAAGCCA 783  
 Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 784 GCAATAAATCTGAATATCTCAGAGCTTTTGAAGGAACTCTATGATTCACCTAGCGCC 843  
 Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 844 CAACGTTACAGTGAATCAGATGGCTGCTTTTACCCCTGCTAAAGATATCATCGTGTACA 903  
 Qy 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 904 CCAATGAGTGGATGCGATCCCGCATGGCCACCATTACCATTATTCCTTACAGCAAGCTT 963  
 Qy 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340  
 Db 964 TCTGCTTTAGAGAAAGATGTCAGATGGCT----- 999  
 Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360  
 Db 999 ----- 999  
 Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380  
 Db 1000 -----ATCAGTGGAACTGGTCTTACAGTCTTCAAGCAATCTCTCTCT 1035  
 Qy 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400  
 Db 1036 AAA----- 1038  
 Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420  
 Db 1039 -----CCTAATGAAGTACTGTCTAGTCTAGGCGAGCTCTTCAAGCAATCTCTCTCT 1089  
 Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440  
 Db 1090 -----TTACGACAAAGTAAAGAGCTCTCTTCAGCATCTGCTGTTATATTTT 1137  
 Qy 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460  
 Db 1138 AITCCAAA---GATATCGTTGAGAAACGGCTACAGCTTATATTTGTAAGACATGCT--- 1191  
 Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480  
 Db 1192 ---GATCACTTCATTACATTCACAAA-----TCAAAATCAAAATGGG 1230  
 Qy 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500  
 Db 1231 CAACCGACTCTCCCAACCAATAGCTAGCAACACCTTCTCCA---TCTCTCCAATCAAT 1287  
 Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520  
 Db 1288 CAGGAACTTCATCAGAAACATGAA----- 1314  
 Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540  
 Db 1315 -----GAAAGATGGATACGGATTTGATGCTAATCGTATATTCGCTGAA 1356  
 Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisIleGlyLysAspSer 560  
 Db 1357 GATGAATCAGGTTTGTTCATGATCAGGAGACCAATCATATTTCTTCAAGAGGAC 1416  
 Qy 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 1417 TTGACAGAGCAAAATTAAGCTGCGCAAAAACATTTAGAGGAA----- 1461  
 Qy 581 ProSerProAspAlaAspValLysAlaAsnProThrGly---AspSerAlaAlaAlaIle 599  
 Db 1462 -----GTTAAACTAGTACATAATGATGATTTCTTCTGTCATCTCAT 1503

Qy 600 TyrAsnArgValLysGly-----GluLys 607  
 Db 1504 GAACAGGATTATCCAGCTAATGCCAAAGAAATGAAGATTAGATAAAAAAATCGAAGAA 1563  
 Qy 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625  
 Db 1564 AAAATCTCGCATATGAAACAATATGGTGTCAACAGTGAAGATATGTCGTAATAAA 1623  
 Qy 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTrp 644  
 Db 1624 GAAAAAATCGATATTTATCCGCATGGAGATCACCATCATGCAGATCCG----- 1674  
 Qy 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664  
 Db 1675 ATTGATGAACAT-----AAACCGTTCGAATGGT-----CATTCCTCAC 1713  
 Qy 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681  
 Db 1714 AGTAACATAGTAACGTGTTTAAACCCGAGAGAGAGTGTCTAAAGAGAGGGAATAAGTT 1773  
 Qy 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 1774 TATACTGGAGAGAAATTAACGAATGTTGTTAATTTGTTAAAAAATAGTACGTTTAATAAT 1833  
 Qy 701 LysAsnPhe 703  
 Db 1834 CAAAACTTT 1842

RESULT 19  
 US-09-884-465A-2  
 ; Sequence 2, Application US/09884465A  
 ; Publication No. US20030077293A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shire Biochem, Inc.  
 ; APPLICANT: Hamel, Josee  
 ; APPLICANT: Brodeur, Bernard  
 ; APPLICANT: Martin, Denis  
 ; APPLICANT: Charland, Nathalie  
 ; APPLICANT: Ouellet, Catherine  
 ; TITLE OF INVENTION: Streptococcus Antigens  
 ; FILE REFERENCE: 055190-0044  
 ; CURRENT APPLICATION NUMBER: US/09/884,465A  
 ; CURRENT FILING DATE: 2001-06-20  
 ; PRIOR APPLICATION NUMBER: 60/212,683  
 ; PRIOR FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 384  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 5048  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-884-465A-2

Alignment Scores:  
 Pred. No.: 9,27e-103 Length: 5048  
 Score: 1247.00 Matches: 298  
 Percent Similarity: 54.36% Conservative: 95  
 Best Local Similarity: 41.22% Mismatches: 180  
 Query Match: 29.94% Indels: 150  
 DB: 10 Gaps: 19

US-09-765-271-56 (1-796) x US-09-884-465A-2 (1-5048)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20  
 Db 1840 GCCTATGCATTAACACGATCGTTCGAGGAAAAAAGACAATAATCGTCTCTTAT 1899  
 Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 1900 GTGGATGGCAGCCAGTCAGTCAAGTCAAGAAAGTGAAGAACTTGACACAGCAGGTAGCCAG 1959  
 Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60

Db 1960 AAAGAGGAATTCAGCGTGAAGCAAAATGTAATCAAAATTTACAGATCAGGGCTATGTAACG 2019

QY 61 SerHisGlyAspHisTyrHisTyrHisGlyValProTyrAspAlaIleIleSer 80

Db 2020 TCACACGGTGACCATATCATTTATATATGGAAGTTCCTTATGATGCCCTCTTTAGT 2079

QY 81 GluGluLeuLeuMetLysAspProHisTyrLysLeuLysAspGluAspIleValAsnGlu 100

Db 2080 GAAGAAGCTCTGATGAAGATCCAACTATCACTTAAAGAGCGTGATATTGTCATGAA 2139

QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrLysAspAla 120

Db 2140 GTCAGAGGTGGTTATCATCAAGTCCGATGAAATATTTATGTCTAGCTGAAGATGCA 2199

QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGluHisSer 140

Db 2200 GCTCATGCTGATATGTTGCAATCAAGATCAATCAATCGTCAAAACCAAGACATGTC 2259

QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160

Db 2260 AAAGATAATGAG-----AAGGTTAACTCTAATGTTGCTGTAGCAAGGTCTCAG 2307

QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180

Db 2308 GGACGATATAGCAAAATGATGTTATGCTTTAATCCAGCTGATATTTATCGAAGATACG 2367

QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200

Db 2368 GGTAATGCTTATATCGTTCTCTCATGGAGTCACTATCACTACATCTCCAAAGCGATTTA 2427

QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220

Db 2428 TCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAATATGCAACCGAGT 2487

QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240

Db 2488 CAGTTAGCTATTCTTCAACAGCTAGTGACAAAT----- 2520

QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260

Db 2521 -----ACACGCAATCTGTAGCAAAAGGATCAACTAGCAAGCCCA 2559

QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuTyrLysLeuProLeuSer 280

Db 2560 GCAAAATAAATCTGAAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTAGGCC 2619

QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProIleGlnIleThrSerArgThr 300

Db 2620 CAACGTTACAGTGAATCAGATGGCCCTGGTCTTTGACCCCTGCTAAGATTATCAGTCGTACA 2679

QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320

Db 2680 CCAAAATGGAGTTGCGATTCGCGATGCGACCATACCATTTATTCCTTACGAAGCTT 2739

QY 321 SerGluLeuGluGluArgIleAlaArgIleProLeuArgTyrArgSerAsnHisTrp 340

Db 2740 TCTGCTTTAGAGAAAGATGTCAGATGGTGCCT----- 2775

QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360

Db 2775 ----- 2775

QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380

Db 2776 -----ATCAGTGAACGTGTTCTACAGTTTCTACAAATGCA 2811

QY 381 ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400

Db 2812 AAA----- 2814

QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420

Db 2815 -----CCTAATGAAGTAGTGTCTAGCTAGGCAAGTCTTTCAAGCAATCCTTCTTCT 2865

QY 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440

Db 2866 -----TTAAACGACAGTAAGGAGCTCTCTTTCAGCATCTGATGGTTATATTTT 2913

QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe\*\*\*AsnLysGlyArg 460

Db 2914 AATCCAAA-----GATATCGTTGAAGAAACGGCTACAGCTTATATTGTAAGACATGGT--- 2967

QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480

Db 2968 ---GATCATTTCCATTATCCAAA-----TCAAAATCAAAATTCGG 3006

QY 481 LysGluLysLeuValAspAspLeuAlaPheLeuAlaProIleThrHisProGluArg 500

Db 3007 CAACCGACTCTTCCAAACATAGTCTAGCAACACCTTCTCCA---TCTCTTCCATCAAT 3063

QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520

Db 3064 CCAGGAACCTTCACATGAGAAACATGAA----- 3090

QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540

Db 3091 -----GAAGATGGATACGATTTGATGCTAATCGTATTTATTCGCTGAA 3132

QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTyrIleGlyLysAspSer 560

Db 3133 GATGAATCAGGTTTGTTCATGAGTCACGAGACCAACATCATTTATTTCTTCAAGAGAC 3192

QY 561 LeuSerAspLysGluLysValAlaAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580

Db 3193 TTGACAGAGAGCAAAATTAAGGCTGCGCAAAACATTTAGAGGAA----- 3237

QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGly---AspSerAlaAlaAlaIle 599

Db 3238 -----GTTAAACATGATGATAATGGATTAGATTCTTTGTCTATCTCAT 3279

QY 600 TyrAsnArgValLysGly-----GluLys 607

Db 3280 GAACAGGATTTATCCAGTAAATGCAAGAAATGAAGATTTAGATAAAAAATCGAAGAA 3339

QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625

Db 3340 AAAATTTGCTGGCAATTATGAACAATATGTTGTCAAAACGTCAAAAGTATTGTCGTAATAA 3399

QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnLysPheAlaTrp 644

Db 3400 GAAAAAATCGCATTTATTTATCCGCATGGAGATCACCATCATCGACATCGG----- 3450

QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664

Db 3451 ATTGATGAACAT-----AAACCGTTGGAATTTGGT-----CATTCAC 3489

QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681

Db 3490 AGTAATGTAACACTGTTTAAACCCGAGAGAGGATTTGCTTAAAAAAGAGGAATAAGTT 3549

QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700

Db 3550 TATATCTGGAGACAATTAACGATTTGTTAATTTGTTAAAAAATAGTAGCTTTAATAAT 3609

QY 701 LysAsnPhe 703

Db 3610 CAAAACCTTT 3618

## RESULT 20

US-10-158-844-192/c

; Sequence 192, Application US/10158844

; Publication No. US20040029118A1

; GENERAL INFORMATION:

; APPLICANT: Kunsch et al.

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391



CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude Pentium 3  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/158,844  
FILING DATE: 03-Jun-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/961,527  
FILING DATE: 1997-10-30  
APPLICATION NUMBER: US 60/029,960  
FILING DATE: 1996-10-31

ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB340PID1

INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6867 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 192:  
US-10-158-844-192

Alignment Scores:  
Pred. No.: 2,396-100 Length: 6867  
Score: 1222.50 Matches: 295  
Percent Similarity: 54.22% Conservative: 97  
Best Local Similarity: 40.80% Mismatches: 181  
Query Match: 29.35% Indels: 151  
DB: 13 Gaps: 18

US-09-765-271-56 (1-796) x US-10-158-844-192 (1-6867)

QY	1	SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr	20
DB	6722	GCCTATGCCTAAACACGATCGTTCGAGAAATAAGCAACAATAATCGTCTCTTAT	6663
QY	21	IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys	40
DB	6662	GTGGATGGCAGCCAGTCAAGTCAAGAAAGTGAAACTTGACACACAGGTTAGCCAG	6603
QY	41	ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr	60
DB	6602	AAAGAGGAAATTCAGCTGAGCAATTTGTAATCAAAATTAACAGATCAGGGCTATGTAACG	6543
QY	61	SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer	80
DB	6542	TCACCGGTGACCATCATCTATTACTATAATGGGAAAGTTCCTTATGATGCCCTCTTTAGT	6483
QY	81	GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu	100
DB	6482	GAAGAACTCTTGATGAAGGATCCAACTATCAACTTAAGACGCTGATATTGTCATGAA	6423
QY	101	ValLysGlyGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla	120
DB	6422	GTCAAGGGTGGTTATATCATCAAGTGCATGGGAAATATTATGTTCTACCTGAAAGATGCA	6363
QY	121	AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnHisSer	140
DB	6362	GCTCATGCTGATAATGTTCGAACATAAGATCAATCGTCAAAACAGAAACATGTC	6303
QY	141	GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln	160

DB	6302	AAAGATAATGAG-----AAGTTAACTCTAAAGTTCTGCTAGCAAGTCTCAG	6255
QY	161	GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr	180
DB	6254	GGAGCATATACGACAAATGATGTTATGCTTTTAAATCCAGCTGATATTATCGAAGATACG	6195
QY	181	GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu	200
DB	6194	GGTAATGCTTATATCGTTCTCATGAGGTCACATCATCATCATCTCCAAAGAGATTTA	6135
QY	201	SerAlaSerGluLeuAlaAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn	220
DB	6134	TCTGCTAGTAGTAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAATAATATCCACCGAGT	6075
QY	221	SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTyrValProSer	240
DB	6074	CAGTTAAGCTTATTTTCAACAGCTAGTGACAAAT-----	6042
QY	241	ValSerAsnProGlyThrThrAsnThrSerSerAsnAsnSerAsnThrAsnSerGln	260
DB	6041	-----AACACGCAATCTCTAGCAAAAGAGATCAACTAGCAAGCCA	6003
QY	261	AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer	280
DB	6002	GCAAATAAATCTGAAATCTCCAGAGTCTTTTGAAGGAACTCTATGATTCACCTAGCGCC	5943
QY	281	GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr	300
DB	5942	CAACGTTACAGTGAATCAGATGGCTGCTTTGACCCCTGCTAAGATTATCAGTCGTACA	5883
QY	301	AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet	320
DB	5882	CCMAATGAGTTGCGATTCGCGATGCGACCATTAACCACTTTATCTTACAGCAAGCTT	5823
QY	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp	340
DB	5822	TCTGCTTTAGAGAAAGATTGCCAGAAATGGTGCCT-----	5787
QY	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
DB	5787	-----	5787
QY	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
DB	5786	-----ATCAGTGGAACTGCTTACAGTTTCTCAAAATGCA	5751
QY	381	ArgLysValGlyGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla	400
DB	5750	AAA-----	5748
QY	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
DB	5747	-----CCTAATGAAGTAGTGTCTAGTCTAGGCAGCTTTTCAAGCAATCTCTCTCT	5697
QY	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr	440
DB	5696	-----TTAACGACAAAGTAAGGAGCTCTCTTCAGCATCTGATGTTATATTTT	5649
QY	441	AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
DB	5648	AAATCCAAAA---GATATCGTTGAAGAAACGGCTACAGCTATATTTGTAAGACATGGT---	5595
QY	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
DB	5594	---GATCAATTTCCATTCATTCACAAA-----TCAAATCAAAATGGG	5556
QY	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
DB	5555	CAACCGCACTCTCCAAACAAATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAAT	5499
QY	501	LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu	520

Db 5498 CCAGGACTTCACATGAGAAACATGAA----- 5472  
 QY 521 AlaAspLysThrThrSerAspGlyTyrIlePheAspGluHisAspIleLeuSerAsp 540  
 Db 5471 -----GAAATGATACGAGTTTGTGCTAATCGTATTATCGCTGAA 5430  
 QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560  
 Db 5429 GATGATCAGGTTTGTGTCATGAGTCAGCGAGACCAACATCATTTCTTCAAGAGGAC 5370  
 QY 561 LeuSerAspLysGluLysValAlaGlnAlaTyrThrLysGluLysGlyIleLeuPro 580  
 Db 5369 TTGACAGAGCAAAATTA--GSTCGCAAAACATTTAGAGGAA----- 5326  
 QY 581 ProSerProAspAlaAspValLysAlaAsnProThrGlyAspSerAlaAlaIleTyr 600  
 Db 5325 -----GTTAAACTAGTCATAATGATTTAGATTCTTTGTCATCTCAT 5284  
 QY 601 AsnArg-----ValLysGlyGluLys 607  
 Db 5283 GAACAGGATTATCCAAAGTAATGCCAAAGAAATGAAGATTAGATAAATAAATCGAGAA 5224  
 QY 608 ArgIleProLeuValArgLeuProTyrMetVal-----GluHisThrValGluValLys 625  
 Db 5223 AAAATGCTGCATTTGAAACAATATGCTGCAACGTAAGAGTATTGCTGTAATAA 5164  
 QY 626 AsnGlyAsnLeuIleIle---ProHisLysAspHisTyrHisAsnIleLysPheAlaTrp 644  
 Db 5163 GAAAAAATGCGATTATTATCCGATGAGATCACCATCATCGATCCG----- 5113  
 QY 645 PheAspAspHisThrTyrLysAlaProAsnGlyTyrThrLeuGluAspLeuPheAlaThr 664  
 Db 5112 ATTGATGAACAT-----AAACCGGTTGCAATTGCT-----CATTCAC 5074  
 QY 665 IleLysTyrTyrValGluHisProAspGluArgProHisSerAsnAspGly----- 681  
 Db 5073 AGTAACTATGACTGTTTAAACCGAAGAGAGCTTGTCTAAAAAGAGGAAATAAGTT 5014  
 QY 682 ---TrpGlyAsnAlaSerGluHisValLeuGlyLysLysAspHisSerGluAspProAsn 700  
 Db 5013 TATACTGGAGAGAAATTAACGAATGTTGTTAAATTTGTTAAATAATAGTCGTTTAATAAT 4954  
 QY 701 LysAsnPhe 703  
 Db 4953 CAAAACTTT 4945  
 RESULT 21  
 US-09-769-787-246  
 ; Sequence 246, Application US/09769787  
 ; Publication No. US20030091577A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Microbial Technics Limited  
 ; APPLICANT: Gilbert, Christophe FG  
 ; APPLICANT: Hansbro, Philip M  
 ; TITLE OF INVENTION: Proteins  
 ; FILE REFERENCE: FWC/P21129WO  
 ; CURRENT APPLICATION NUMBER: US/09/769,787  
 ; PRIOR FILING DATE: 2001-01-26  
 ; PRIOR APPLICATION NUMBER: GB 9816337.1  
 ; PRIOR FILING DATE: 1998-03-27  
 ; PRIOR APPLICATION NUMBER: US 60/125164  
 ; PRIOR FILING DATE: 1999-03-19  
 ; NUMBER OF SEQ ID NOS: 388  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 246  
 ; LENGTH: 1455  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-769-787-246  
 Alignment Scores:  
 Pred. No.: 1,51e-99 Length: 1455  
 Score: 1203.50 Matches: 259

Percent Similarity: 59.47% Conservative: 77  
 Best Local Similarity: 45.84% Mismatches: 120  
 Query Match: 28.90% Indels: 109  
 DB: 10 Gaps: 10

US-09-765-271-56 (1-796) x US-09-769-787-246 (1-1455)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnArgValSerTyr 20  
 Db 64 GCCTATGCACCTAAACCCAGCATCGTTCCGAGGAAATAGGACAAATATCGTCTCTTAT 123  
 QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40  
 Db 124 GTGGATGCGACCCAGTCAGTCAAGTCAGAAAAAGTGAACACTTGACACCAAGCAGGTAGCCAG 183  
 QY 41 ArgGluGlyIleAsnAlaGluLysIleValIleLysIleThrAspGlnGlyTyrValThr 60  
 Db 184 AAAGAGGAATTCAGGCTGAGCAAAATTTGTAATCAAAATTACAGATCAGGGGTATGTAACG 243  
 QY 61 SerHisGlyAspHisTyrHisTyrTyrAsnGlyLysValProTyrAspAlaIleSer 80  
 Db 244 TCACACGGTGACCACTATCATTAATAATGSGAAAGTTCCTTATGATGCCCTCTTTAGT 303  
 QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLysLysAspGluAspIleValAsnGlu 100  
 Db 304 GAAGAACTCTTGATGAAGGATCCAAACTATCAACTTAAAGACGCTGATATTGTCAATGAA 363  
 QY 101 ValLysGlyTyrValIleLysValAspGlyLysTyrTyrValTyrLeuLysAspAla 120  
 Db 364 GTCAAGGGTGGTTATATCATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 423  
 QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140  
 Db 424 GCTCATGCTGATAATGTTGGAACATAAGATGAAATCAATCGTCAAAACACAGAACATGTC 483  
 QY 141 GlnHisArgGluGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160  
 Db 484 AAAGATAATCAG-----AAGGTTAACTCTAATGTTGCTGAGCAAGTCTCAG 531  
 QY 161 GlyArgTyrThrThrAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180  
 Db 532 GGACGATATACGACAAATGATGTTATGCTTTAATCCAGCTGATATTATCGAAGATACG 591  
 QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200  
 Db 592 GGTAATGCTTATATCGTCTCATGAGGTCACATCACTACATCTCCAAAAGCGATTGA 651  
 QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220  
 Db 652 TCTGCTAGTGAATTAGCAGCGCTAAAGCACATCTCGCTGGAATAAATATGCAACCGAGT 711  
 QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240  
 Db 712 CAGTTAAGCTATCTTCAACAGCTAGTGCAAT----- 744  
 QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260  
 Db 745 -----AACCGCAATCTGTAGCAAAAGGATCACTAGCAAGCCCA 783  
 QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280  
 Db 784 GCAATAAATCTGAAAAATCTCCAGAGTCTTTTGAAGGAACTCTATGATTCACCTAGCGCC 843  
 QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300  
 Db 844 CAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 903  
 QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320  
 Db 904 CCAAAATGAGTGGATTCCGATCGCGACCACTTACCACTTTATTCCTTACGCAAGCTT 963  
 QY 321 SerGluLeuGluArgIleAlaArgIleLeuProLeuArgTyrArgSerAsnHisTrp 340

```

Db 964 TCTGCCTTAGAAGAAAGATTGCCAGNATGGTGCTT----- 999
Qy 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db 999 ----- 999
Qy 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db 1000 -----ATCAGTGGAACTCGTTCTACAGCTTCTCAAAATGCA 1035
Qy 381 ArgLysValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAla 400
Db 1036 AAA----- 1038
Qy 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db 1039 -----CCTAATGAAGTAGTGTCTAGTCTAGCGAGCTTCTCAAGCAATCTCTCTTCT 1089
Qy 421 ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db 1090 -----TTAACGACAAAGTAAGAGCTCTCTCAGCATCTGATGGTATATTTT 1137
Qy 441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe**AsnLysGlyArg 460
Db 1138 AATCCAAA---GATATCGTTGAAGAAACCGCTACAGCTTATTTGTAAGACATGGT--- 1191
Qy 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db 1192 ---GATCTTCCATTCATTCCAAA-----TCAAAATCAAAATGGG 1230
Qy 481 LysGluLysLeuValAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db 1231 CAACCGACTCTTCCAAACAATAGTCTAGCAACACCTCTCTCCA---TCTCTTCCATCAAT 1287
Qy 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db 1288 CAGGAACCTTCATCAGAAACATGAA----- 1314
Qy 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db 1315 -----GAGATGATACCGATTGATGCTAAATCGTATATTCGCTGAA 1356
Qy 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisIleIleGlyLysAspSer 560
Db 1357 GATGAATCAGGTTTGTCTCAGAGTCCAGGAGACCAATCATATTTCTTCAGAGAGGAC 1416
Qy 561 LeuSerAspLysGlu 565
Db 1417 TTGACAGAGAGCAA 1431

```

RESULT 22

```

US-09-769-744A-23
; Sequence 23, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1455

```

```

; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-23

Alignment Scores: 1.51e-99 Length: 1455
Pred. No.: 1203.50 Matches: 259
Score: 59.47% Conserved: 77
Best Similarity: 45.84% Mismatches: 120
Best Local Similarity: 28.90% Indels: 109
Query Match: 10 Gaps: 10
DB:

US-09-765-271-56 (1-796) x US-09-769-744A-23 (1-1455)

Qy 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
Db 64 GCCTATGCATTAACACGAGCATCGTTCAGGAAATAAGGCAATAATCGTGTCTCTTAT 123
Qy 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
Db 124 GTGGATGGCAGCCAGTCAAGTCAGAAAGTGAAACCTTGACACGAGACCGGCTAGCCAG 183
Qy 41 ArgGluGlyIleAsnAlaGluGlnIleValIleLysIleThrAspGlnGlyTyrValThr 60
Db 184 AAGAAGGAATTCAGGCTGACCAATTCGTAATCAAAATTCAGATCAGGGCTATGTAACG 243
Qy 61 SerHisGlyAspHisTyrTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
Db 244 TCACACGGTGACCACTATCATTACTATAATGGAAAGTTCCTTATGATGCCCTCTTAGT 303
Qy 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysAspGluAspIleValAsnGlu 100
Db 304 GAAGAACTCTTGATGAGGATCCAACTATCAACTTAAAGACGCTGATATGTCATGAA 363
Qy 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
Db 364 GTCAAGGGTGGTTATATCATCAAGTCGATGGAATATATTATGCTACCTGAAGATGCA 423
Qy 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
Db 424 GCTCATGCTGATATATGTTTCGAACTAAAGATGAATCAATCGTCAAAAACAAGAATGTC 483
Qy 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
Db 484 AAAGATAATGAG-----AAGGTAACTCTAATGTTGCTGTACCAAGGCTCAG 531
Qy 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
Db 532 GGACGATATACGACAAATGATGTTATGCTTTTATCCAGCTGATATTATCGAAGATACG 591
Qy 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
Db 592 GGTAAATGCTTATATCTCTCTCATGGAGGTCATCTACTATACATCTCCAAAAGCGATTTA 651
Qy 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
Db 652 TCTGTAGTAGTAATTAGCAGCAGCTAAAGCACATCTCGCTGGAAAAAATATGCAACCGAGT 711
Qy 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db 712 CAGTTAAGCTATTCTTCAACAGCTAGTGACAAI----- 744
Qy 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db 745 -----AACACGCAATCTGTAGCAAAAGATCACTAGCAACCCA 783
Qy 261 AlaSerGlnSerAsnAspIleAspSerLeuLysGlnLeuLysLeuTyrLysLeuProLeuSer 280
Db 784 GCAAAATAATCTGAAATCTCCAGAGTCTTTTGAGGAACCTCTATGATTACCTTAGCGCC 843
Qy 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db 844 CAACGTTACAGTGAATCAGATGGCTCTTTGACCCCTGCTAAGATATATCAGTCGTACA 903

```

Qy	301	AlaArgGlyValAlaValProHisGlyAspHisTyHisPheIleSerProTySerGlnMet	320
Db	904	CCAAATGGAGTTGCGATTCCGATCGCGACCATTAACCATTTATTCCTTACGCAAGCTT	963
Qy	321	SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyArgSerAsnHisTrp	340
Db	964	TCTGCCTTAGAAGAAAGATTCCCAAGATTGGTGCCT	999
Qy	341	ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly	360
Db	999	-----	999
Qy	361	ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal	380
Db	1000	-----ATCAGTGAAGTGGTTCTACAGTTTCTTACAAATGCA	1035
Qy	381	ArgLysValGlyGluGlyTyValPheGluGluLysGlyIleSerArgTyValPheAla	400
Db	1036	AAA-----	1038
Qy	401	LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer	420
Db	1039	-----CCTAATGAAGTAGTGTCTAGTCAGCAGCTCTTCAAGCAATCCTCTTCT	1089
Qy	421	ValSerHisThrLeuThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTy	440
Db	1090	-----TTAACGACAAAGTAAGGAGCTCTCTCAGCATCTGATGGTTATATTTT	1137
Qy	441	AspLysAlaTyAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg	460
Db	1138	AATCCAAAA--GATATCGTGTGAAGAACGGCTACAGCTTATATGTGAAGACATGCT--	1191
Qy	461	AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn	480
Db	1192	---GATCATTTCCATTACATTCACAA-----TCAATCAAAATTCGG	1230
Qy	481	LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg	500
Db	1231	CAACCGACTCTTCCAAACAAATAGTCTAGCAACACCTTCTCCA--TCTCTTCCAATCAAT	1287
Qy	501	LeuGlyLysProAsnSerGlnIleGluTyThrGluAspGluValArgIleAlaGlnLeu	520
Db	1288	CCAGGAACCTTCACATGAGAAACATGAA-----	1314
Qy	521	AlaAspLysTyThrThrSerAspGlyTyIlePheAspGluHisAspIleIleSerAsp	540
Db	1315	-----GAAGATGGATACGGATTTCATGCTAATCGTATATTCGCTGAA	1356
Qy	541	GluGlyAspAlaTyValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer	560
Db	1357	GATGATCATGGTTTCTCATGCTAGCTACGGAGACCAACATCATATTATTTCTTCAAGAGGAC	1416
Qy	561	LeuSerAspLysGlu	565
Db	1417	TTGACAGAAGACAA	1431

RESULT 23

RESULTS 23  
US-10-412-862-7

CS 10-412-882- /  
; Sequence 7, Application US/10412862

Sequence 7, Application US/10412781A1  
; Publication No. US20040052781A1

REGISTRATION NO. 0320  
; GENERAL INFORMATION:

APPLICANT: Johnson, Leslie S.

APPLICANT: Adamou, John E.

;  
TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

**TITLE OF INVENTION:** Pneumoniae Polypeptides Having Selected Structural

; TITLE OF INVENTION: Motifs

FILE REFERENCE: 469201-685

; CURRENT APPLICATION NUMBER: US/10/412,862

; CURRENT FILING DATE: 2003-04-14

; PRIOR APPLICATION NUMBER: 09/468,656  
 ; PRIOR FILING DATE: 1000-10-01

; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 507.

; PRIOR APPLICATION NUMBER: 60/113,048

```

784 GCAATTAATCTGAAATCTCCAGAGCTTTTGAAGGAACTCTATGATTCACCTAGCGCC 843
281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
844 CAAGTTACAGTGAATCAGATGCGTCTTTGACCCCTGCTAAGATTTATCAGTCGTACA 903
301 AlaArgGlyValAlaValProHisGlyAspHisThrPheIleProTyrSerGlnMet 320
904 CCAATCGAGTGGATGCGCATGCGCATGGGACCAITACCACCTTTATCTTACAGCAAGCTT 963
321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
964 TCTCCCTTAGAGAAAGATGCGCAGATGGTCT----- 999
341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
999 ----- 999
361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
1000 -----ATCAGTGGAACTGGTCTACAGTCTTCTACAAATGCA 1035
381 ArgLysValGlyGluGlyTyrValPheGluLysGlyIleSerArgTyrValPheAla 400
1036 AAA----- 1038
401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
1039 -----CCTAATGAAGTAGTGTCTAGCTAGCAGCTCTTCAAGCAATCCTTCTTCT 1089
421 ValSerHisThrLeuAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
1090 -----TTAACGACAAAGTAGGAGCTCTCTCAGCATCTGATGTTATATTTT 1137
441 AspLysAlaTyrAsnLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
1138 AATCAAAA--GATATCGTTGAAGAACGGCTACAGCTTATATTGTAAGACATGCT--- 1191
461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
1192 ---GATCAATTCCATTACATTCACAAA-----TCAATCAATGGG 1230
481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
1231 CAACCGACTCTTCCAAACAATAGCTAGCAACACCTCTCCA---TCTCTTCAATCAAT 1287
501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
1288 CCAGGAACTTCACATGAGAAACATGAA----- 1314
521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
1315 -----GAAGATGGATACGGATTTGATGCTAATCGTATATTCGCTGAA 1356
541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
1357 GATGATCAGGTTTGTGTCATGAGTCACGGAGACCAATCATTTATTCITCAAGAGGAC 1416
561 LeuSerAspLysGlu 565
1417 TTGACAGAGAGCA 1431

```

```

RESULT 24
US-10-412-850-7
; Sequence 7, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-696

```

```

; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 7
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-10-412-850-7

Alignment Scores:
Pred. No.: 1,51e-99 Length: 1455
Score: 1203.50 Matches: 259
Percent Similarity: 59.47% Conservative: 77
Best local Similarity: 45.84% Mismatches: 120
Query Match: 28.90% Indels: 109
DB: 16 Gaps: 10

US-09-765-271-56 (1-796) x US-10-412-850-7 (1-1455)

QY 1 SerTyrGluLeuGlyLeuTyrGlnAlaArgThrValLysGluAsnAsnArgValSerTyr 20
DB 64 GCCTATGACCTAAACACCCAGCATCGTTCCGAGGAAATAAGGACAATATCGTGCTCTTAT 123
QY 21 IleAspGlyLysGlnAlaThrGlnLysThrGluAsnLeuThrProAspGluValSerLys 40
DB 124 GTGATGGCAGCCAGTCAGTCAGTCAGAAAGTGAACCTTGACACACAGCAGGTTAGCCAG 183
QY 41 ArgGluGlyIleAsnAlaGluGlnIleValLysIleThrAspGlnGlyTyrValThr 60
DB 184 AAAGAGGAATTCAGGCTGAGCAAAATGTAATCAAAATTTACAGATCAGGGCTATGTACG 243
QY 61 SerHisGlyAspHisTyrHisTyrAsnGlyLysValProTyrAspAlaIleIleSer 80
DB 244 TCACACGGTGACCATCTATCTATCTATTAATGGGAAAGTTCTTATGATGCCCTCTTAGT 303
QY 81 GluGluLeuLeuMetLysAspProAsnTyrLysLeuLysLysAspGluAspIleValAsnGlu 100
DB 304 GAAGAACTCTTGATGAGGATCCAACTATCAACTTAAAGCGCTGATATGTCAATGAA 363
QY 101 ValLysGlyGlyTyrValLysValAspGlyLysTyrTyrValTyrValLysAspAla 120
DB 364 GTCAAGGGTGGTTATATCATCATCAAGTCGATGGAATAATATTATCTCTACCTGAAAGATGCA 423
QY 121 AlaHisAlaAspAsnValArgThrLysGluGluIleAsnArgGlnLysGlnGluHisSer 140
DB 424 GCTCATGCTGATAATGTTCCGAATGAAGATGAATCAATCAATCGTCNAAACCAAGACATGTC 483
QY 141 GlnHisArgGluGlyGlyThrProArgAsnAspGlyAlaValAlaLeuAlaArgSerGln 160
DB 484 AAAGATAATGAG-----AAGGTTAACTCTAATGTTGCTAGCAAGGCTCTCAG 531
QY 161 GlyArgTyrThrThrAspAspGlyTyrIlePheAsnAlaSerAspIleIleGluAspThr 180
DB 532 GGACGATATACGACAAATGATGTTATGTTTAAATCCAGCTGATATTCGAAGATACG 591
QY 181 GlyAspAlaTyrIleValProHisGlyAspHisTyrHisTyrIleProLysAsnGluLeu 200
DB 592 GGTAAATGCTTATATCTGTTCTCTCTGAGGAGTCATCATCATCATCCCAAAAGCATTTA 651
QY 201 SerAlaSerGluLeuAlaAlaGluAlaPheLeuSerGlyArgGlyAsnLeuSerAsn 220
DB 652 TCTGCTAGTGAATTAGCAGCAGCTAAAGCAGCATCTGGCTGGAATAATATGCAACCGAGT 711
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
DB 712 CAGTTAAGCTATTCTTCAACAGCAGTAGTGACAAAT----- 744
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260

```



Fri Oct 1 15:35:52 2004

Search completed: October 1, 2004, 13:50:16  
Job time : 1107 secs

```
QY 221 SerArgThrTyrArgArgGlnAsnSerAspAsnThrSerArgThrAsnTrpValProSer 240
Db :|||
712 CAGTTAAGCTATTCTTCAACAGCTAGTGACAAT----- 744
QY 241 ValSerAsnProGlyThrThrAsnThrAsnThrSerAsnAsnSerAsnThrAsnSerGln 260
Db :|||
745 -----AACCCCAATCTGTAGCAAAAGGATCACTAGCAAGCCA 783
QY 261 AlaSerGlnSerAsnAspIleAspSerLeuLeuLysGlnLeuTyrLysLeuProLeuSer 280
Db :|||
784 GCAATAAATCTGAATCTCCAGAGCTTTTGAAGGAACCTCTATGATTCACCTTAGCGGCC 843
QY 281 GlnArgHisValGluSerAspGlyLeuValPheAspProAlaGlnIleThrSerArgThr 300
Db :|||
844 CAACGTTACAGTGAATCAGATGGCTGCTCTTCCACCTGTAAAGATTATCAGTCGTACA 903
QY 301 AlaArgGlyValAlaValProHisGlyAspHisTyrHisPheIleProTyrSerGlnMet 320
Db :|||
904 CCAATGGAGTGGATTCGCGATGGCGACCATTACCACCTTATTCCTTACAGCAAGCTT 963
QY 321 SerGluLeuGluGluArgIleAlaArgIleIleProLeuArgTyrArgSerAsnHisTrp 340
Db :|||
964 TCTGCCCTTAGAAGAAAAGATTCCAGAAATGGTGCT----- 999
QY 341 ValProAspSerArgProGluGlnProSerProGlnProThrProGluProSerProGly 360
Db :|||
999 ----- 999
QY 361 ProGlnProAlaProAsnLeuLysIleAspSerAsnSerSerLeuValSerGlnLeuVal 380
Db :|||
1000 -----ATCAGTGAAGCTGGTTCTCAGAGTTCTCACAATGCA 1035
QY 381 ArgLysValGlyGluGlyTyrValPheGluGluLysGlyIleSerArgTyrValPheAla 400
Db :|||
1036 AAA----- 1038
QY 401 LysAspLeuProSerGluThrValLysAsnLeuGluSerLysLeuSerLysGlnGluSer 420
Db :|||
1039 -----CCTAATGAAGTAGTGTCTAGCTAGGAGTCTTCAAGCAATCCTTCTTCT 1089
QY 421 ValSerHisThrThrAlaLysLysGluAsnValAlaProArgAspGlnGluPheTyr 440
Db :|||
1090 -----TTAAGCAGAGTAAGAGCTCTCTTCAGCATCTGATGTTATATTTT 1137
QY 441 AspLysAlaTyrAsnLeuLeuThrGluAlaHisLysAlaLeuPhe***AsnLysGlyArg 460
Db :|||
1138 AATCCAAAA---GATATCGTTGAAGAAACGGCTACAGCTTATATCTAAGACATGGT--- 1191
QY 461 AsnSerAspPheGlnAlaLeuAspLysLeuLeuGluArgLeuAsnAspGluSerThrAsn 480
Db :|||
1192 ---GATCATTTTCATTACATTCAAAA-----TCAATCAAAATGGG 1230
QY 481 LysGluLysLeuValAspAspLeuLeuAlaPheLeuAlaProIleThrHisProGluArg 500
Db :|||
1231 CAAACCACTCTCCAAACATAGTCTAGCAACACCTTCTCCA---TCTCTTCCAATCAAT 1287
QY 501 LeuGlyLysProAsnSerGlnIleGluTyrThrGluAspGluValArgIleAlaGlnLeu 520
Db :|||
1288 CCAGGAACCTTCACATGAGAAACATGAA----- 1314
QY 521 AlaAspLysTyrThrThrSerAspGlyTyrIlePheAspGluHisAspIleIleSerAsp 540
Db :|||
1315 -----GAAGATGGATACGGATTGATGCTAATCGTATTATCGCTGAA 1356
QY 541 GluGlyAspAlaTyrValThrProHisMetGlyHisSerHisTrpIleGlyLysAspSer 560
Db :|||
1357 GATCAATCAGGTTTGTGTCATGAGTCACGGAGACCAACATCATTTATTTCTTCAAGAGGAC 1416
QY 561 LeuSerAspLysGlu 565
Db :|||
1417 TTGACAGAGAGACAA 1431
```



90 7 0.9 104 17 Q8u029 pyrococcus

## ALIGNMENTS

## RESULT 1

Q9AHT9  
ID Q9AHT9 PRELIMINARY; PRT; 816 AA.  
AC Q9AHT9;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Pneumococcal histidine triad A protein.  
GN PHTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus  
CX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tucumanet E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.,  
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules  
Affording Protection Against Streptococcus pneumoniae Infection."  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL: AF291695; AAK:9155.1; -  
DR InterPro: IPR006270; Strep\_his\_triad.  
DR Pfam: PF04270; strep\_his\_triad; 5.  
DR TIGRFAMs: TIGR01363; strep\_his\_triad; 2.  
SQ SEQUENCE 8-6 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 99.9%; Score 795; DB 2; Length 816;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAQIVIKITDQGYVT 60  
DB 21 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAQIVIKITDQGYVT 80  
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYIVKDGKYYVYLKDA 120  
DB 81 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYIVKDGKYYVYLKDA 140  
QY 121 AHADNVRTKEEINRQKCHSQHREGGTPRNDGAVALARSGRYTTDGYIFNASDIIEDT 180  
DB 141 AHADNVRTKEEINRQKCHSQHREGGTPRNDGAVALARSGRYTTDGYIFNASDIIEDT 200  
QY 181 GDAYIVPHGDHYHYIPKNELASLAAAEAFLSGRGNLSNRRYRQNSDNTSRTNWVPS 240  
DB 201 GDAYIVPHGDHYHYIPKNELASLAAAEAFLSGRGNLSNRRYRQNSDNTSRTNWVPS 260  
QY 241 VSNPGTNTNTSNNTNSQASQSDNDISLLKOLYKPLSQRHVESDGLVDPAPQITSRT 300  
DB 261 VSNPGTNTNTSNNTNSQASQSDNDISLLKOLYKPLSQRHVESDGLVDPAPQITSRT 320  
QY 301 ARGVAVPHGDHYHYIPYSQMSLEERIEARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 360  
DB 321 ARGVAVPHGDHYHYIPYSQMSLEERIEARIIPLYRSNHWVDSRPEQSPQPTPEPSPG 380  
QY 361 PQAPNLKIDSSLSVQLVRKVGEGYFEKGISRYVFAKOLPSETVKNLESKLSKQBS 420  
DB 381 PQAPNLKIDSSLSVQLVRKVGEGYFEKGISRYVFAKOLPSETVKNLESKLSKQBS 440  
QY 421 VSHLTATKENVAPRQEPYDAIYNLLTEAHKALFNKGRNSDFCALDKLLERLNDESTN 480  
DB 441 VSHLTATKENVAPRQEPYDAIYNLLTEAHKALFNKGRNSDFCALDKLLERLNDESTN 500  
QY 481 KEKLVDDLLAFLAPITHPERLKGKPSQIEYTEDVRIAQLADKYITSDGYIFDEHDIISD 540

DB 501 KEKLVDDLLAFLAPITHPERLKGKPSQIEYTEDVRIAQLADKYITSDGYIFDEHDIISD 560  
QY 541 EGDAYVTPHGHSHWIGKDSLSKEKVAAQAYTKEGILPSPDADVKANFTGSAAYI 600  
DB 561 EGDAYVTPHGHSHWIGKDSLSKEKVAAQAYTKEGILPSPDADVKANFTGSAAYI 620  
QY 601 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIIPKOHYHNIKEAFWDDHTYKAPNGYITLED 660  
DB 621 NRVKGEKRIPLVRLPYMVEHTVEVNGNLIIIPKOHYHNIKEAFWDDHTYKAPNGYITLED 680  
QY 661 LFTATIKYVVEHPDERPHSNDGNGWGNASEHVLGKDHSDPNKFNKADBEPEETPAEVEVP 720  
DB 681 LFTATIKYVVEHPDERPHSNDGNGWGNASEHVLGKDHSDPNKFNKADBEPEETPAEVEVP 740  
QY 721 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLNRLTIQIMDNNSIMAEAEKLLA 780  
DB 741 QVETEKVEAQLKEAEVLLAKVTDSSLSKANATETLAGLNRLTIQIMDNNSIMAEAEKLLA 800  
QY 781 LLKGSNPSSVSKEKIN 796  
DB 801 LLKGSNPSSVSKEKIN 816

## RESULT 2

Q97QM8  
ID Q97QM8 PRELIMINARY; PRT; 802 AA.  
AC Q97QM8;  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Conserved domain protein.  
GN SP1175.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
CX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC BAA-334 / TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson M.C., Peterson C.D.,  
RA Mayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson I., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL: AE007418; AAK75284.1; -  
DR PIR: C95136; C95136.  
DR TIGR: SP1175; -  
DR InterPro: IPR006270; Strep\_his\_triad.  
DR Pfam: PF04270; strep\_his\_triad; 5.  
DR TIGRFAMs: TIGR01363; strep\_his\_triad; 2.  
KW Complete proteome.  
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1933 CRC64;

Query Match 87.3%; Score 695; DB 16; Length 802;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAQIVIKITDQGYVT 60  
DB 7 SYELGLYQARTVKNRRVSYIDGKQATOKTENLTPEVSKREGINAQIVIKITDQGYVT 66  
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYIVKDGKYYVYLKDA 120  
DB 67 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDKEDIIVNEVKGYIVKDGKYYVYLKDA 126

**This Page Blank (uspto)**

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 07:24:16 ; Search time 78 Seconds  
(without alignments)  
3219.904 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 796

Sequence: 1 SYELGYQARTYKNNRVSY.....KLLALLKGSNPSVSKKIN 796

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 10-7041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 90 summaries

Database :

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mmc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirts.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	99.9	816	2	Q9AHT9
2	695	87.3	802	16	Q97QM8
3	616	77.4	828	16	Q8DPQ2
4	107	13.4	844	2	Q9AG74
5	107	13.4	855	16	Q8CWR4
6	60	7.5	819	2	Q9ANY3
7	60	7.5	819	16	Q97QM9
8	60	7.5	819	16	Q9ANY2
9	60	7.5	853	16	Q8DQ08
10	32	4.0	1033	16	Q9ANY1
11	32	4.0	1039	16	Q8DQ07
12	27	3.4	481	16	Q8E338
13	17	2.1	289	2	Q9AE21
14	17	2.1	794	16	Q877Y2
15	17	2.1	822	2	Q9ZHG7
16	17	2.1	822	16	Q8E4U1

17	17	2.1	822	16	Q8D281	Q8d281 streptococc
18	17	2.1	823	16	Q8N282	Q8n282 streptococc
19	17	2.1	823	16	Q8K5Q1	Q8k5q1 streptococc
20	17	2.1	825	2	Q93GT5	Q93gt5 streptococc
21	17	2.1	825	16	Q99XV4	Q99xv4 streptococc
22	17	2.1	182	16	Q8DQ06	Q8dq06 streptococc
23	10	1.3	250	16	Q7WNG8	Q7wng8 bordetella
24	10	1.3	250	16	Q7WB03	Q7wb03 bordetella
25	10	1.3	250	16	Q7VY29	Q7vy29 bordetella
26	9	1.1	791	16	Q9GL01	Q9gl01 pasteurallia
27	8	1.0	71	12	Q8B9M0	Q8b9m0 rachiplusia
28	8	1.0	117	16	Q89NM4	Q89nm4 bradyrhizob
29	8	1.0	147	16	Q8EHD8	Q8ehd8 shewanella
30	8	1.0	152	5	Q8IIB9	Q8iib9 plasmodium
31	8	1.0	154	5	Q91742	Q91742 hydra atten
32	8	1.0	181	5	Q44095	Q44095 drosophila
33	8	1.0	270	5	Q9NCF0	Q9ncf0 cryptocercu
34	8	1.0	284	5	Q22866	Q22866 caenorhabdi
35	8	1.0	284	5	Q22865	Q22865 caenorhabdi
36	8	1.0	284	5	Q27284	Q27284 caenorhabdi
37	8	1.0	294	5	Q46010	Q46010 caenorhabdi
38	8	1.0	301	5	Q95Q31	Q95q31 caenorhabdi
39	8	1.0	319	16	Q8FLF7	Q8flf7 corynebacte
40	8	1.0	330	13	Q7ZWH1	Q7zwh1 brachydancio
41	8	1.0	400	16	Q82AV2	Q82av2 streptomyc
42	8	1.0	436	3	Q9P8N1	Q9p8n1 coriolus ve
43	8	1.0	484	16	Q9KS51	Q9ks51 oryza sativ
44	8	1.0	485	10	Q80401	Q80401 oryza sativ
45	8	1.0	488	16	Q99TD8	Q99td8 staphylococ
46	8	1.0	488	16	Q8W40	Q8w40 staphylococ
47	8	1.0	490	16	Q8CNR2	Q8cnr2 staphylococ
48	8	1.0	673	6	Q95LH9	Q95lh9 oryctolagus
49	8	1.0	675	16	Q89MC4	Q89mc4 bradyrhizob
50	8	1.0	840	16	Q9RSJ4	Q9rsj4 helicobacte
51	8	1.0	945	16	Q26077	Q26077 helicobacte
52	8	1.0	1169	2	Q84EVA	Q84ev4 methylococc
53	8	1.0	1205	16	Q7TXM2	Q7txm2 mycobacteri
54	8	1.0	1233	17	Q8TMX4	Q8tmx4 methanosarc
55	8	1.0	1646	2	Q93HW0	Q93hw0 streptococc
56	8	1.0	1659	16	Q97S90	Q97s90 streptococc
57	8	1.0	1659	16	Q8C252	Q8c252 streptococc
58	8	1.0	2621	5	Q8SSW5	Q8ssw5 dictyosteli
59	7	0.9	42	15	Q87936	Q87936 chimpanzee
60	7	0.9	42	15	Q87948	Q87948 chimpanzee
61	7	0.9	42	15	Q87938	Q87938 chimpanzee
62	7	0.9	42	15	Q88084	Q88084 chimpanzee
63	7	0.9	42	15	Q87952	Q87952 chimpanzee
64	7	0.9	42	15	Q87946	Q87946 chimpanzee
65	7	0.9	42	15	Q88080	Q88080 chimpanzee
66	7	0.9	43	15	Q87944	Q87944 chimpanzee
67	7	0.9	44	15	Q87942	Q87942 chimpanzee
68	7	0.9	49	15	Q87940	Q87940 chimpanzee
69	7	0.9	49	16	Q45596	Q45596 bacillus su
70	7	0.9	50	2	Q9AGM9	Q9agm9 clostridium
71	7	0.9	51	16	Q8PHV0	Q8phv0 escherichia
72	7	0.9	53	10	Q7XK3	Q7xk3 oryza sativ
73	7	0.9	68	15	Q90VA0	Q90va0 human immu
74	7	0.9	68	16	Q9A3B0	Q9a3b0 caulobacter
75	7	0.9	71	1	Q977L2	Q977l2 uncultured
76	7	0.9	71	17	Q96YD8	Q96yd8 sulfolobus
77	7	0.9	80	3	Q9HGR7	Q9hgr7 gilbeistella
78	7	0.9	81	16	Q83L13	Q83l13 shigella fl
79	7	0.9	83	5	Q86ID1	Q86id1 dictyosteli
80	7	0.9	86	15	Q73850	Q73850 human immu
81	7	0.9	86	15	Q73851	Q73851 human immu
82	7	0.9	86	15	Q73852	Q73852 human immu
83	7	0.9	90	17	Q97W98	Q97w98 sulfolobus
84	7	0.9	92	16	Q83KC6	Q83kc6 shigella fl
85	7	0.9	93	16	Q9RUG5	Q9rug5 deinococcus
86	7	0.9	94	16	Q9RY27	Q9ry27 deinococcus
87	7	0.9	96	16	Q8FAP6	Q8fap6 escherichia
88	7	0.9	98	2	Q7WV63	Q7wv63 bacillus su
89	7	0.9	98	16	Q7V2U9	Q7v2u9 prochloroco

**This Page Blank (uspto)**

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=211357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Swinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radzue D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter C.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001);
DR EMBL; AF318956; AAK06761.1;
DR EMBL; AE007403; AAK75121.1;
DR PIR; H95115; H95115;
DR TIGR; SP1004;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 6;
DR TIGRFAMs; TIGR01363; strep_his_triad; 4;
KW Signal, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
SQ SEQUENCE 1039 AA; 11463 MW; 81A563FC806625C4 CRC64;

Query Match 4.0%; Score 32; DB 16; Length 1039;
Best Local Similarity 100.0%; Pred. No. 5.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
Db 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98

RESULT 11
Q8DQ07 PRELIMINARY; PRT; 1039 AA.
AC Q8DQ07; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein E.
GN PHE OR SP0908
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6)
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.F., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008454; AAK99712.1;
DR PIR; D97985; D97985;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 6;
DR TIGRFAMs; TIGR01363; strep_his_triad; 3;
KW Complete proteome.
SQ SEQUENCE 1039 AA; 114625 MW; 05CC226D2028F551 CRC64;

Query Match 4.0%; Score 32; DB 16; Length 1039;
Best Local Similarity 100.0%; Pred. No. 5.9e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 46 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 77
Db 67 AEQIVIKITDQGYVTSHGHDHYHYNGKVPYDA 98

RESULT 12
Q8E338 PRELIMINARY; PRT; 481 AA.
AC Q8E338;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN GHS1925;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766854; CAD47584.1;
DR Sagaliet; G81925;
DR InterPro; IPR06270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 4;
DR TIGRFAMs; TIGR01363; strep_his_triad; 1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 481 AA; 53325 MW; DEF0453D8A929BE CRC64;

Query Match 3.4%; Score 27; DB 16; Length 481;
Best Local Similarity 100.0%; Pred. No. 2.5e-17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 GYVTSHGHDHYHYNGKVPYDAIISEEL 83
Db 81 GYVTSHGHDHYHYNGKVPYDAIISEEL 107

RESULT 13
Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
CX NCBI_TaxID=1311;
RN [-]
RP SEQUENCE FROM N.A.
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Mielch F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBS11, an active group
RT II intron identified in human isolates of group B streptococci.";
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL; AJ290952; CAC35985.1;
DR InterPro; IPR006270; Strep_his_triad;
DR Pfam; PF04270; strep_his_triad; 2;
DR TIGRFAMs; TIGR01363; strep_his_triad; 1;
KW Hypothetical protein.
FT NON TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;

```

**This Page Blank (uspto)**

---

```

RT pneumoniae";
RL Science 293:498-506(2001);
DR EXML; AEC07418; AAK75283.1; -.
DR PIR; B95136; B95136.
DR TIGR; SP1174; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E9163BDE CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 819;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 8
Q9ANY2 PRELIMINARY; PRT; 839 AA.
AC Q9ANY2
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Dornitz M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
RX SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguilo S., Dickinson T., Hickey B.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06760.1; -.
DR EMBL; AEC07403; AAK75120.1; -.
DR PIR; G95115; G95115.
DR TIGR; SP1003; -.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW SIGNAL, Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 16; Length 839;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 9
Q8DQ08 PRELIMINARY; PRT; 853 AA.
AC Q8DQ08
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Pneumococcal histidine triad protein D.
GN PHTD OR SP9097.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.S., Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.P., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsuoka P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicase T.L.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R., Jr., Skatrud P.L.,
RA Glas J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99711.
DR PIR; C97985; C97985.
DR InterPro; IPR006270; Strep_his_triad.
DR Pfam; PF04270; strep_his_triad; 5.
DR TIGRFAMs; TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MW; 9BF06A1EDE9590CF5 CRC64;

Query Match
Best Local Similarity 7.5%; Score 60; DB 15; Length 853;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

31 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 90
52 ENLTPDEVSKREGINAEQIVIKITDQGYTSHGDHYYNGKVPYDAIISELLMKDPNY 111

RESULT 10
Q9ANY1 PRELIMINARY; PRT; 1039 AA.
AC Q9ANY1
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004).
GN PHT E OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.B., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dornitz M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).

```



**This Page Blank (uspto)**

---



**This Page Blank (uspto)**

---

OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=1313;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2124685; PubMed=11349048;  
 RA Zhang Y., Masi A.W., Barniak V., Mountzuros K., Hostetter M.K.,  
 RA Green B.A.;  
 RT "Recombinant PnpA Protein, a Unique Histidine Motif-Containing Protein  
 RT from Streptococcus pneumoniae, Protects Mice against Intranasal  
 RL Pneumococcal Challenge";  
 DR EMBL; AF340221; AAK26629.1; -  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad; 5  
 DR TIGRFAMs; TIGR01363; strep\_his\_triad; 2.  
 SQ SEQUENCE 844 AA; 94769 MW; 5738A52290FF8902 CRC64;  
 Query Match 13.4%; Score 107; DB 2; Length 844;  
 Best Local Similarity 130.0%; Pred. No. 7.1e-96;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 KGGVIVKDGKYYVYLKDAHADNVRTKEENRQKQHSQHREGTPRNDGAVALARSG 161  
 DB 122 KGGVIVKDGKYYVYLKDAHADNVRTKEENRQKQHSQHREGTPRNDGAVALARSG 181  
 QY 162 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 208  
 DB 182 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 228  
 RESULT 5  
 Q8CWR4  
 ID Q8CWR4 PRELIMINARY; PRT; 855 AA.  
 AC Q8CWR4;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DE Histidine motif-containing protein.  
 GN PHPA OR SP1060.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=171101;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee J.N., Leikowitz E.J., Lu J., Matsushima P.,  
 RA McAhren S.M., McMenney M., McLeaster K., Mundy C.W., Niclas T.I.,  
 RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR ENML; AB008479; AAK99864.1; -  
 DR PIR; D98004; D98004.  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad; 5.  
 DR TIGRFAMs; TIGR01363; strep\_his\_triad; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 855 AA; 96177 MW; 4350B82A3F97089A CRC64;  
 Query Match 13.4%; Score 107; DB 16; Length 855;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-96;  
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 102 KGGVIVKDGKYYVYLKDAHADNVRTKEENRQKQHSQHREGTPRNDGAVALARSG 161  
 DB 122 KGGVIVKDGKYYVYLKDAHADNVRTKEENRQKQHSQHREGTPRNDGAVALARSG 181  
 QY 162 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 208  
 DB 182 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 228

DB 133 KGGVIVKDGKYYVYLKDAHADNVRTKEENRQKQHSQHREGTPRNDGAVALARSG 192  
 QY 162 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 208  
 DB 193 RYTTDDGYIENASDIETGDAYIVPHGDHYYHPKNSLASELAAA 239  
 RESULT 6  
 Q9ANY3  
 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.  
 AC Q9ANY3;  
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Pneumococcal histidine triad protein B precursor (fragment).  
 GN PHTB.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=1313;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21101045; PubMed=11159990;  
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,  
 RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,  
 RA Langermann S., Koenig S., Johnson S.,  
 RT "Identification and Characterization of a novel family of pneumococcal  
 RT proteins (the Pht family) that are protective against sepsis";  
 RL Infect. Immun. 69:949-958(2001).  
 DR EMBL; AF318954; AAK06759.1; -  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad; 5.  
 DR TIGRFAMs; TIGR01363; strep\_his\_triad; 2.  
 KW SIGNAL  
 FT NON\_TER 819 29 POTENTIAL.  
 SQ SEQUENCE 819 AA; 92109 MW; E602CFC16CC28A5F CRC64;  
 Query Match 7.5%; Score 60; DB 2; Length 819;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-49;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 31 ENLTDEVSKREGINAEQIVIKITDQGYVTSAGDHYHYNGKVPYDAIISLLMKDPNY 90  
 DB 52 ENLTDEVSKREGINAEQIVIKITDQGYVTSAGDHYHYNGKVPYDAIISLLMKDPNY 111  
 RESULT 7  
 Q97QM9  
 ID Q97QM9 PRELIMINARY; PRT; 819 AA.  
 AC Q97QM9;  
 DT 01-OCT-2001 (T-EMBLrel. 18, Created)  
 DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Conserved domain protein.  
 GN SP1174.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 RN NCBI\_TaxID=1313;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC BAA-334 / TIGR4;  
 MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.S.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Hoit I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus

**This Page Blank (uspto)**

---

Query Match 2.1%; Score 17; DB 2; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85  
 |||||  
 DB 92 YNGKVPYDAIISELLM 108

RESULT 14

Q877Y2 ID Q877Y2 PRELIMINARY; PRT; 794 AA.  
 AC Q877Y2;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN SPS1722.  
 OS Streptococcus pyogenes (serotype M3).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=198466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SSI-1 / Serotype M3;  
 RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yamashita A.,  
 RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hattori M.,  
 RA Hayashi H., Hamada S.;  
 RT "The genome of invasive Streptococcus pyogenes; a comparative analysis  
 of S. pyogenes SSI-1, SF370 and MGAS8232."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP005146; BAC64817.1;  
 DR InterPro: IPR006270; Strep\_his\_triad.  
 DR Pfam: PF04270; strep\_his\_triad; 6.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 794 AA; 9807269C8F5CB875 CRC64;

Query Match 2.1%; Score 17; DB 16; Length 794;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85  
 |||||  
 DB 63 YNGKVPYDAIISELLM 79

RESULT 15

Q9ZHG7 ID Q9ZHG7 PRELIMINARY; PRT; 822 AA.  
 AC Q9ZHG7;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Streptococcus agalactiae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1311;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R268;  
 RX MEDLINE=99115568; PubMed=9916102;  
 RA Spellerberg B., Rodzinski E., Martin S., Weber-Heynemann J.,  
 RA Schnitzler N., Luetticken R., Podbielski A.;  
 RT "Imb, a protein with similarities to the Irai adhesin family, mediates  
 attachment of Streptococcus agalactiae to human laminin."  
 RL Infect. Immun. 67:811-818(1999).  
 DR EMBL: AF062533; AAD13797.1;  
 DR FIR; T46758; T46758.  
 DR InterPro: IPR006270; Strep\_his\_triad.  
 DR Pfam: PF04270; strep\_his\_triad; 6.  
 DR TIGRFAMS: TIGR01363; strep\_his\_triad; 4.

KW Hypothetical protein.  
 SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313491F98 CRC64;

Query Match 2.1%; Score 17; DB 2; Length 822;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85  
 |||||  
 DB 92 YNGKVPYDAIISELLM 108

RESULT 16

Q8E4U1 ID Q8E4U1 PRELIMINARY; PRT; 822 AA.  
 AC Q8E4U1;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN GBS1306.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Maadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease."  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 DR EMBL: AL766850; CAD46965.1;  
 DR Sagalish; gbs1306;  
 DR InterPro: IPR006270; Strep\_his\_triad.  
 DR Pfam: PF04270; strep\_his\_triad; 6.  
 DR TIGRFAMS: TIGR01363; strep\_his\_triad; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 822 AA; 92393 MW; 2929A97C8AFCD78F CRC64;

Query Match 2.1%; Score 17; DB 16; Length 822;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 YNGKVPYDAIISELLM 85  
 |||||  
 DB 92 YNGKVPYDAIISELLM 108

RESULT 17

Q8DZ81 ID Q8DZ81 PRELIMINARY; PRT; 822 AA.  
 AC Q8DZ81;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Streptococcal histidine triad family protein.  
 GN SAG1233.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222988; PubMed=12200547;  
 RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,

RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,  
RA Cary H.A., Cline K.T., Van Aken S.E., Gll J., Scarcelli M., Mora M.,  
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,  
RA Rinaudo D., Rappucoli R., Telford J.L., Kasper D.L., Grandi G.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative genomic analysis of an  
RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
RL EMBL; AE014248; AAN00111.1; -;  
DR TIGR; SAG1233; -;  
DR InterPro; IPR006270; Strep\_his\_triad.  
DR Pfam; PF04270; strep\_his\_triad; 6.  
DR TIGRFAMs; TIGR01363; strep\_his\_triad; 1.  
KW Complete proteome.  
SQ SEQUENCE 822 AA; 92400 MW; 8CC8DF316727F98 CRC64;  
Query Match 2.1%; Score 17; DB 16; Length 822;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YNGKVPYDAIISEELM 85  
|||||  
Db 92 YNGKVPYDAIISEELM 108  
RESULT 18  
Q8N282 PRELIMINARY; PRT; 823 AA.  
AC Q8N282;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein spym18\_2072.  
GN SPYM18\_2072.  
OS Streptococcus pyogenes (serotype M18).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=186103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS232 / Serotype M18;  
RX MEDLINE=21927593; PubMed=11917108;  
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;  
RT "Genome sequence and comparative microarray analysis of serotype M18  
RT group A Streptococcus strains associated with acute rheumatic fever  
RT outbreaks.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
DR EMBL; AE010110; AAL98543.1; -;  
DR InterPro; IPR006270; Strep\_his\_triad.  
DR Pfam; PF04270; strep\_his\_triad; 6.  
DR TIGRFAMs; TIGR01363; strep\_his\_triad; 4.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 823 AA; 92585 MW; C7951EB30CEEDP0C CRC64;  
Query Match 2.1%; Score 17; DB 16; Length 823;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YNGKVPYDAIISEELM 85  
|||||  
Db 92 YNGKVPYDAIISEELM 108  
RESULT 19  
Q8K5Q1 PRELIMINARY; PRT; 823 AA.  
AC Q8K5Q1;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Histidine triad protein.  
GN SPYM3\_1724.  
OS Streptococcus pyogenes (serotype M3).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=198466;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGAS315 / Serotype M3;  
RX MEDLINE=22133808; PubMed=12122206;  
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,  
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,  
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,  
RA Schlievert P.M., Musser J.M.;  
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:  
RT phage-encoded toxins, the high-virulence phenotype, and clone  
RT emergence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).  
RL EMBL; AE014169; AAM80331.1; -;  
DR InterPro; IPR006270; Strep\_his\_triad.  
DR Pfam; PF04270; strep\_his\_triad; 6.  
DR TIGRFAMs; TIGR01363; strep\_his\_triad; 1.  
KW Complete proteome.  
SQ SEQUENCE 823 AA; 92431 MW; 8C6CBD517A2DD616 CRC64;  
Query Match 2.1%; Score 17; DB 16; Length 823;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YNGKVPYDAIISEELM 85  
|||||  
Db 92 YNGKVPYDAIISEELM 108  
RESULT 20  
Q93GT5 PRELIMINARY; PRT; 825 AA.  
AC Q93GT5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Histidine triad protein of group A streptococci.  
GN HTPA.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SSI-9;  
RA Terao Y., Kawabata S., Hamada S.;  
RT "Characterization of a novel histidine triad protein of group A  
RT streptococci.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AB073859; BAB71774.1; -;  
DR InterPro; IPR006270; Strep\_his\_triad.  
DR Pfam; PF04270; strep\_his\_triad; 6.  
DR TIGRFAMs; TIGR01363; strep\_his\_triad; 4.  
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;  
Query Match 2.1%; Score 17; DB 2; Length 825;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 69 YNGKVPYDAIISEELM 85  
|||||  
Db 92 YNGKVPYDAIISEELM 108  
RESULT 21  
Q99XV4 PRELIMINARY; PRT; 825 AA.  
ID Q99XV4  
AC Q99XV4;



DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Hypothetical protein SPY2006.  
 GN SPY2006  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192884; PubMed=11296296;  
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 DR EMBL; AB006623; AAK34688.1; -;  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad\_6  
 DR TIGRFAMs; TIGR01363; strep\_his\_triad; 4.  
 KW Hypothetical protein; Complete proteome.  
 SQ HYPOTHETICAL 825 AA; 92649 MW; 57DF50969E6D50F4 CRC64;  
  
 Query Match 2.1%; Score 17; DB 16; Length 825;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-07; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0;  
  
 Qy 69 YNGKVPYDAIISEELM 85  
 Db 92 YNGKVPYDAIISEELM 108  
 |||||  
  
 RESULT 22  
 Q8DQ06 PRELIMINARY; PRT; 182 AA.  
 AC Q8DQ06;  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Pneumococcal histidine triad protein E, truncation.  
 GN PHEE-TRUNCATION OR SP0910.  
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 RA Gilmore R., Glass J.S., Knoja H., Kraft A.R., Lagace R.E.,  
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushina P.,  
 RA McAlhnen S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,  
 RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,  
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL; AE008464; AAK9714.1; -;  
 DR PIR; F97985; F97985.  
 DR InterPro; IPR006270; Strep\_his\_triad.  
 DR Pfam; PF04270; strep\_his\_triad\_2.  
 KW Complete proteome.  
 SQ SEQUENCE 182 AA; 20145 MW; 29E12D653313A343 CRC64;  
  
 Query Match 1.3%; Score 10; DB 16; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 0.58;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 188 HGDHYHYIPK 197  
 Db 35 HGDHYHYIPK 44  
 |||||  
  
 RESULT 23  
 Q7WNG8 PRELIMINARY; PRT; 250 AA.  
 AC Q7WNG8;  
 DT 01-OCT-2003 (TREMELrel. 25, Created)  
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Putative ATP-binding component of ABC transporter.  
 GN BB1423.  
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,  
 RA Cerdono-Tarraga A.M., Temple L., James K., Basham D., Mungall K.L.,  
 RA Achtman M., Atkin R., Baker S., Temple L., James K., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,  
 RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 DR EMBL; BX640441; CAE31921.1; -;  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 250 AA; 27826 MW; 4D6D89723660A7F5 CRC64;  
  
 Query Match 1.3%; Score 10; DB 16; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 0.76;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 515 VRTAQLADKY 524  
 Db 133 VRTAQLADKY 142  
 |||||  
  
 RESULT 24  
 Q7WB03 PRELIMINARY; PRT; 250 AA.  
 AC Q7WB03;  
 DT 01-OCT-2003 (TREMELrel. 25, Created)  
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Putative ATP-binding component of ABC transporter.  
 GN BP1206.  
 OS Bordetella parapertussis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12822 / ATCC BAA-587;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,  
 RA Cerdono-Tarraga A.M., Temple L., James K., Basham D., Mungall K.L.,  
 RA Achtman M., Atkin R., Baker S., Temple L., James K., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,  
 RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,  
RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
RT Bordetella parapertussis and Bordetella bronchiseptica.",  
RL Nat. Genet. 35:32-40(2003).  
DR EMBL: BX640426; CAE36508.1; --  
KW ATP-binding; Complete proteome.  
SQ SEQUENCE 250 AA; 27826 MW; 4D6D89723660A7F5 CRC64;

Query Match 1.3%; Score 10; DB 16; Length 250;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 VRIAGLADKY 524  
| | | | | | | | | |  
Db 133 VRIAGLADKY 142

RESULT 25  
Q7VY29 PRELIMINARY; PRT; 250 AA.  
AC Q7VY29;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative ATP-binding component of ABC transporter.  
GN BP1534.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;  
RX MEDLINE=22827954; PubMed=12910271;  
RA Parkhill J., Sebatia M., Preston A., Murphy L.D., Thomson N.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,  
RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
RT Bordetella parapertussis and Bordetella bronchiseptica.",  
RL Nat. Genet. 35:32-40(2003).  
DR EMBL: BX640415; CAE41823.1; --  
KW ATP-binding; Complete proteome.  
SQ SEQUENCE 250 AA; 27928 MW; 3EDAFD80C5BECCCA CRC64;

Query Match 1.3%; Score 10; DB 16; Length 250;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 515 VRIAGLADKY 524  
| | | | | | | | | |  
Db 133 VRIAGLADKY 142